Pricing Policy for Development Management

EDITED BY
Gerald M. Meier
Pricing Policy
for
Development Management

EDI SERIES IN ECONOMIC DEVELOPMENT
Pricing Policy
for
Development Management

Edited by
Gerald M. Meier
Preface vii

Introduction 1

   1.1 The Technological Choices Open to Any Society
       Paul A. Samuelson 5
   1.2 A Picture of Prices and Markets
       Paul A. Samuelson 13
   1.3 Elements of Demand and Supply
       Paul A. Samuelson 17
   1.4 Elasticity of Demand and Supply
       Paul A. Samuelson 23
   1.5 Applications of Supply and Demand
       Milton H. Spencer 29
   1.6 What Do We Mean by “Cost”? Milton H. Spencer 36
   1.7 The Economic Organization of a P.O.W. Camp
       R. A. Radford 39
   1.8 The Nature and Necessity of a Price System
       J. A. Schumpeter 42
   1.9 Economic Responsiveness in Developing Economies
       Peter T. Bauer and Basil S. Yamey 44
   1.10 The Market Mechanism as an Instrument of Development
       Harry G. Johnson 45
   1.11 Prices in a Socialist Economy
       Oscar Lange and Fred M. Taylor 49
   1.12 The Model of Choice
       Edith Stokey and Richard Zeckhauser 54

2. Functions of Prices 65
   2.1 Limited Knowledge and Economic Analysis
       Kenneth Arrow 65
   2.2 Market and Planning
       Janos Kornai 66
   2.3 Distortions of Agricultural Incentives
       Gilbert T. Brown 73
   2.4 Pricing of Public Urban Services
       Irving A. Sirken 77
   2.5 Congestion Pricing—The Example of Singapore
       Peter L. Watson and Edward P. Holland 80
   2.6 Congestion Levies and Surcharges for Ports
       Esra Bennathan and A. A. Walters 82
   2.7 Price Responsiveness in Agriculture
       Raj Krishna 84
   2.8 Interest Rates on Agricultural Loans
       Gordon Donald 90
   2.9 Financial Deepening
       Edward S. Shaw 95
   2.10 Implications of Factor Choice and Technological Change
       in Transport
       Vincent W. Hogg 98
   2.11 Agricultural Pricing Policies and Income Distribution
       Gilbert T. Brown 100
   2.12 Agricultural Prices, Distribution, and Employment
       John Mellor 102
   2.13 Consequences of Farm Tractors in Pakistan
       Graham Donaldson 104
   2.14 Who Benefits from Government Expenditure: Colombia
       Marcelo Selowsky 110

3. Properties of Efficiency 114
   3.1 Efficiency of the Firm: Short Period
       Paul A. Samuelson 114
3.2 Implications for Economic Welfare  Paul A. Samuelson  136
3.3 The Fragmented Economy  R. I. McKinnon  138
3.4 The Nature of Restricted Competition  Robert Dorfman  140
3.5 Economic Inefficiency  Tibor Scitovsky  145
3.6 Transfer Pricing and Developing Countries  Sanjaya Lall  146
3.7 Commodity Markets and Commodity Agreements  Jere R. Behrman  148
3.8 The Problem of Social Cost  R. H. Coase  154
3.9 Externalities in Developing Countries  I. M. D. Little and J. A. Mirrles  159
3.10 The Nature of Public Goods  Peter O. Steiner  164

4. Public Pricing Policy  170
4.1 Principles of Public Pricing  Richard M. Bird  171
4.2 Marginal Cost Pricing in Practice: The Case of a Bridge  Ray Rees  176
4.3 Marginal Cost Pricing  Marcel Boiteux  179
4.4 Marginal Cost Pricing for Electricity Undertakings  Arnold C. Harberger  183
4.5 Electricity Tariff Policy  Ralph Turvey and Dennis Anderson  184
4.6 Energy Pricing Policy  D. G. Fallen-Bailey and T. A. Byer  190
4.7 Cost of Using Roads  A. A. Walters  192
4.8 Port Pricing  Esra Bennathan and A. A. Walters  199
4.9 Alternative Definitions of Marginal Cost  Robert J. Saunders, Jeremy J. Warford, and Patrick C. Mann  203
4.10 Water Supply Pricing  Robert J. Saunders and Jeremy J. Warford  207
4.11 Public Transport Pricing and Cost Recovery  A. R. Prest  216
4.12 Evaluating Fertilizer Subsidies  Dana G. Dalrymple  222
4.13 The Willingness-to-Pay Criterion  Ralph Turvey and Dennis Anderson  227

5. Prices, Markets, and Development  231
5.1 The Price Mechanists  Frances Stewart and Paul Streeten  234
5.2 Choice of Policy Instruments  Hollis B. Chenery  236
5.3 Cost-Benefit Analysis in Developing Countries  F. Leslie C. H. Helmers  242
5.4 Some Preliminary Theory  Deepak Lal  247
5.5 Nature of Project Analysis  Lyn Squire and Herman G. van der Tak  249

Further Reading  255
Index  259
Preface

This book is for the development practitioner who wants to learn more about the role of prices in a country's development process. Although it presupposes no formal training in economics, it does assume that the reader wishes to understand the essential elements of a price system, the functions of prices, the various policies that a government might pursue in cases of market failure, and the principles of public pricing of goods and services provided by government enterprises. For a student in more specialized development courses, this book may also offer some broad background reading. Although the discussion stops short of expounding the techniques of project appraisal, it is intended to provide a would-be project evaluator with an appreciation of the underlying logical structure of cost-benefit project appraisal. For one interested in more general aspects of policy analysis (whether monetary policy, fiscal policy, or foreign trade policy), a knowledge of price analysis should also help in the selection of policy instruments that will relax the constraints on the development process.

The discussion proceeds on three levels—theory, application, and policy. It would, however, be a false issue to confront theory versus practice, or theory versus policy. These are not dichotomies: they are interrelated. Theory, practice, and policy all contribute to each other. The purpose of public policymaking is to reduce the divergence between a preferred situation and the actual situation. But before a remedial policy can be initiated, theory is first necessary to determine what the actual situation is. The facts do not speak for themselves; one has to integrate them logically through the insights provided by theory. We must be able to ask questions of the facts, give meaning to the facts, and avoid the superficiality of measurement without theory. Although theory is most usefully viewed as the handmaiden of applied economics, a competent applied economist needs sound grounding in theory. And without the capacity to apply economic analysis, it is impossible to make the first necessary step toward policy prescription. After diagnosing the actual situation, a policy advisor must go on to elucidate for the decisionmaker the alternative courses of action available to remedy the actual situation and approach the preferred situation.

To integrate the theoretical, applied, and policy dimensions of price analysis, I have combined a commentary with selected readings that elucidate the basic principles of pricing and illustrate their application to development problems. To give as much concrete substance as possible to the applied and policy dimensions, I have drawn many of the readings from the experience of development practitioners. These relate to the important fields of agriculture, industry, power, urban services, foreign trade, and employment.

Although I hope that the whole of the book is greater than the sum of its parts, a reader may select, of course, portions to study according to individual needs and interests.

I wish to express my appreciation to the authors and publishers who have granted permission to use excerpts from books, articles, and other publications.
for which American or foreign copyrights exist. Specific acknowledgment is given with each selection. Some parts of the original versions of the selected materials have been silently omitted out of consideration for relevancy and the avoidance of repetition. Tables and diagrams have been renumbered; some footnotes have been omitted and others have been renumbered; references to other parts of original sources have been deleted when necessary.

Beyond being grateful to the World Bank as a source for many of the readings, I also appreciate the hospitality and superb facilities provided by the Economic Development Institute of the World Bank during the writing period. The director, Ajit Mozoomdar, was encouraging and most helpful. A special acknowledgment is due J. Price Gittinger, coordinator of training materials and publications at EDI, for first suggesting the subject of this book and then leaving me completely free to write as I wished. The teaching staff of EDI generously shared their experience in offering invaluable advice. Especially helpful were F. Leslie C. H. Helmers and Eugene R. Schlesinger. Charles Magnus, also of EDI, provided invaluable assistance in preparing the manuscript for publication. To the many others who contributed to this undertaking, I am grateful; but the usual disclaimer is made with respect to anyone else's responsibility for the final outcome.

GERALD M. MEIER
Many economists believe that a major task of development policy is to "get prices right." Policy prescriptions frequently emphasize the need for "more appropriate pricing systems" or "a sensible pricing system." But how does one know when prices are "right," "appropriate," or "sensible"? What exactly are the policy instruments that policymakers can use to get prices right? This book explores these questions. In so doing, it emphasizes that pricing has fundamental economic policy implications and is not merely an activity of accountants or financial analysts.

Advice on price policy is as old as economics is a discipline. In his celebrated book, *The Wealth of Nations* (1776), Adam Smith portrayed the price-market system as if it were an "invisible hand," which led every individual, via the forces of competition, to contribute to the general welfare. "Every individual endeavors to employ his capital so that its produce may be of greatest value. He generally neither intends to promote the public interest, nor knows how much he is promoting it. He intends only his own security, only his own gain. And he is in this led by an Invisible Hand to promote an end which was not part of his intention. By pursuing his own interest he frequently promotes that of society more effectually than when he really intends to promote it." The invisible hand thus assures the harmonization of private interest and public interest.

To celebrate the bicentenary of *The Wealth of Nations*, the economics department of the University of Chicago produced a film about Smith—but they had extreme difficulty in trying to illustrate the invisible hand! In this book—directed to the Poverty of Nations—we hope to be more successful in showing the application of pricing policy in concrete situations that typify policy problems confronted by the development manager. In some of these situations, Smith's view may still prevail; but in others, there will be cases of market failure, and remedial policies by the government will be necessary "to promote the public interest."

To recognize whether the invisible hand operates or is absent, and to be able to assess whether private interest does or does not coincide with the public interest, we must first understand some central concepts in economics. Of major importance are the following concepts that will continually recur throughout this book:

*Responsiveness to prices*: The responsiveness of buyers and sellers to price changes—in the amount bought and supplied—must be emphasized. Only too often is it wrongly assumed that quantities bought or sold are unrelated to prices. An early warning by a pioneering development economist must still be repeated: "Much of the current discussion on underdeveloped countries is vitiated by treating supply and demand as physical quantities unaffected by

---

prices, and, perhaps more generally, by regarding quantities as fixed or given, and not variables. This widespread practice strikes at the root of economics as a distinct discipline and threatens to transform it into a branch of production engineering.  

Substitution: The logic of choice runs throughout price analysis. The role of prices is to serve as appropriate substitution ratios in a whole range of choice problems. This involves the substitution of one factor of production for another in production, one product for another in consumption, one policy objective for another in policy formation.  

Mutual interdependence: A change in one market may lead to a change in another market.

Prices are like boats on a smooth lake: not a movement of a single boat, however slight, but starts a ripple which will sooner or later affect every other boat. It is only because they do form parts of a web of interpenetrating influences that prices can perform their function of attracting each particular kind of goods towards those who are willing to pay most for it; of attracting factors of production towards the making of those kinds of goods for which there are the strongest demands; and of balancing the various pulls and thrusts of desire and cost so that, in a broad sense, and given the fulfillment of certain essential conditions which we shall see later, resources are made to go as far as they can towards satisfying people’s needs.  

Efficiency: The properties of technical efficiency must be recognized in the sense of maximizing production. Beyond that, the properties of economic efficiency must be recognized in the sense of optimizing production.  

Economic welfare: The notion of “public interest” must be explored. A distinction between social benefit and private benefit may be necessary. Criteria for improvement or betterment need to be established. A comparison of policy situations is necessary, and policies need to be ranked according to their consequences for economic welfare.

The foregoing concepts enter into the underlying principles of project appraisal. In determining how to make production decisions for social ends, a project evaluator will necessarily incorporate these concepts in the appraisal process. They are implicit in even the most mechanical and rule-of-thumb type of manual for project appraisal. The better they are understood, the more meaningful will be project appraisal.

The practice of project appraisal also requires an understanding of the constraints and policies that determine a country’s development. A project evaluator must therefore be able to interpret how prices affect the constraints on development and how a range of development policies involve price changes. Project appraisal needs to be interpreted within this broader context of the effects of alternative pricing policies on the development process.

In most of the less developed countries, three types of constraints can be singled out for special emphasis: the savings gap, the foreign exchange gap, and the agricultural bottleneck.

Additional private saving, government saving, and foreign saving must finance domestic investment. As long as the volume of savings is limited, so too will be investment and the growth in output. The savings constraint reflects the need
to mobilize a surplus of resources and to invest the resources productively. This process of capital accumulation is composed of three steps: the increase in the volume of real savings—the release of resources for investment purposes; the channeling of savings through a finance and credit mechanism—the claiming of the resources by investors; and the act of investment itself—the use of the resources for increasing the capital stock. Each of these steps can be affected by prices.

The rate of growth in gross national product (GNP) will also be limited by the country's capacity to import. To the extent that there is a foreign exchange gap, the import requirements of a country's development program cannot be fulfilled and development will be retarded. Prices may affect the size of their foreign exchange gap by influencing export revenue, import requirements, and the inflow of foreign capital.

The importance of agricultural development has been increasingly recognized as necessary for the success of industrial development. If the agricultural sector develops too slowly it can be a brake on the rest of the economy by not supplying sufficient food and raw materials to the industrial sector and by not providing a widening market for industrial products. The need for intersectoral balance between the agricultural and industrial sectors is now emphasized to ensure an investible surplus of savings and taxes from the rural sector, a contribution to the balance of payments through exports or possibly import replacement, and the provision of expanded employment opportunities and more productive employment in agricultural activities. If the agricultural constraint is to be relaxed, proper pricing policies in agriculture will be required.

Beyond a relaxation of the savings, foreign exchange, and agricultural constraints, there is also a growing concern over the complex problems of unemployment and underemployment, inequality, and absolute poverty. Despite creditable performance in raising GNP per capita, most of the developing countries realize that other ways to reduce poverty need to be considered as well: improving income distribution, increasing employment, and fulfilling basic needs. The problems of population growth and rapid urbanization also intensify the complexity of development tasks.

To fulfill these changing objectives and deal with these complicated problems, more attention needs to be given to pricing policy in the various dimensions of development management.

If an evaluation of the effects of alternative pricing policies requires an understanding of the principles of pricing, so too does the practice of project appraisal. Cost-benefit analysis of projects has become widespread in developing countries. There is a need for such analysis when the private profitability of a project is not equivalent to its social (or national) profitability because private receipts do not adequately measure social benefits, and private expenditures are not equal to social costs.

The principles outlined in this book should therefore be relevant to a host of development problems—ranging from policies designed to mobilize savings to the techniques of appraising the "newer type" of projects in health or education.

To indicate how the principles of pricing might be applied to these various problems, the discussion is structured in the following way. The first chapter, "Elements of a Price System," explains the components of a price system and relates them to the logic of optimization in a choice problem. Chapter 2, "Functions of Prices," distinguishes the various functions performed by prices and illustrates some consequences of pricing policies. Chapter 3, "Properties of Effi-
ciency," presents an ideal reference model of an efficient economy, so that actual cases of market failure can be compared with the optimal. It also examines how deviations from the ideal market arise and what remedial policies are possible to correct the distortions. Chapter 4, "Public Pricing Policy," outlines contending principles of ideal pricing and considers applications to a range of public services. Chapter 5, "Prices, Markets, and Development," acts as a summary and places the techniques of project appraisal in the wider context of price analysis and investment policy.
Central to economic analysis is the role of prices. Perhaps the most useful knowledge that economics can convey to the development manager is an understanding of the means by which prices allocate resources and mobilize resources. Whether they are determined in markets free from government intervention, or established by a central planning agency, or modified by governmental policies, prices will have considerable influence on any country's rate and pattern of development.

This chapter assembles the elements of a price system and relates prices to the general economic problem of the need to exercise choice whenever scarcity exists. The logic of economic reasoning is the logic of rational choice. But choice implies valuation. Thus, the logic of the price system is very much the logic of optimization in a choice problem. When economic analysis deals with the allocation of scarce resources to alternative uses, it necessarily involves choice and valuation. If there were no scarce resources, all goods would be free goods, and there would be no problem of how to economize on the use of economic (scarce) goods. Or if there were only one use for resources, there would be no problem of how to allocate the resources to alternative ends. The combination of limited resources and alternative uses requires a price system to allocate resources.

The following selection (1.1) underscores the fundamental law of scarcity.

### 1.1 The Technological Choices
Open to Any Society

We have discussed the basic economic fact that limitation of the total resources capable of producing different commodities necessitates a choice between relatively scarce commodities. This can be illustrated quantitatively by simple arithmetic examples and geometrical diagrams.

Consider an economy with only so many people, so much technical knowledge, so many factories and tools, and so much land, water power, and natural resources. In deciding WHAT shall be produced and HOW, the economy must really decide just how these resources are to be allocated among the thousands of different possible commodities. How much land should go into wheat growing? Or into pasturage? How many factories are to produce knives? How much skilled labor for machine shops?

These problems are complicated even to discuss, much less solve. Therefore we must simplify. So let us assume that only two economic goods (or classes of economic goods) are to be produced. For dramatic purposes, we can choose the pair Adolf Hitler ranted about—guns and butter. These two commodities are commonly used to illustrate the problem of choosing between civilian and war goods, but the same analysis applies to any choice of goods. Thus the more resources the government uses to build public roads, the less will be left to produce private houses; the more the public chooses to consume of food, the less it can consume of clothing; the more society decides to consume today, the less can be its production of machines and capital goods to turn out more consumption goods for the next year or decade.

**NUMERICAL EXAMPLE**

But let us stick to the example of guns and butter given in table 1-1 and figure 1-1. If all resources are
Table 1-1. Alternative Production Possibilities of Guns and Butter

<table>
<thead>
<tr>
<th>Possibilities</th>
<th>Butter (millions of pounds)</th>
<th>Guns (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

thrown into the production of civilian goods (butter), there will still be a maximum amount of butter that can be produced per year. (The exact amount depends upon the quantitative and qualitative resources of the economy in question and the technological efficiency with which they are used.) Suppose 5 million pounds of butter is the maximum amount that can be produced with the existing technology and resources.

At the other extreme, imagine that 100 percent of society's resources had been devoted instead to the production of guns. Only some maximum number of guns could then be produced: 15 (thousand) guns of a certain description can perhaps be produced if we are really willing to produce no butter.

These are two extreme possibilities; in between there are still others. If we are willing to give up some butter, we can have some guns. If we are willing to give up still more butter, we can have still more guns. A schedule of a number of possibilities is given in table 1-1, F being the extreme where all butter and no guns are produced, and A being the opposite extreme where all resources go into guns. In between, at E, D, C, and B, butter is being given up increasingly in return for more guns. Butter is "transformed" into guns, not physically, but by diverting resources from one use to the other.

GRAPHICAL PRODUCTION-POSSIBILITY FRONTIER

It is even more illuminating to represent this same production-possibility schedule graphically in figure 1-1, by measuring butter along the horizontal axis and guns along the vertical.

The reader should now be able to go directly from the numerical table to the diagram: for F, by counting over 5 butter units to the right and going up 0 gun units; for E, by going 4 butter units to the right and going up 5 gun units; and finally, for A, by going over 0 butter units and up 15 gun units.

We may fill in all intermediate positions with new dots, even those involving fractions of a million pounds or fractions of a thousand guns—as in the so-called production-possibility frontier shown in figure 1-2.

The curve that we now have represents this fundamental fact:

A full-employment economy must always in producing one good be giving up something of another. Substitution is the law of life in a full-employment economy. The production-possibility frontier depicts society's menu of choices.

UNEMPLOYMENT AND INEFFECTIVENESS

But what if there had been widespread unemployment of resources: idle men, idle land, idle factories? We have already warned that our economic laws may then be quite different. This is one such instance.

With unemployment, we are not on the production-possibility frontier at all, but rather, somewhere inside it. Thus, U in figure 1-2 represents a point inside the p-p frontier where society is producing only 2 (million) pounds of butter and 6 (thousand) guns. If resources are idle, by putting them to work we can have more butter and more guns. We can move from U to D and thereby get more butter and more guns.

Business-cycle unemployment is not the only way of being inside the p-p frontier. If an economy is
Some Uses of the Production-Possibility Frontier

This concept, represented as a simple curve, can help introduce many of the most basic concepts in economics. For example, figure 1·2 illustrates the basic definition of economics, namely, the problem of choosing among scarce or limited resources ("means" capable of alternative uses), in order to achieve best goals ("ends"). Land, labor, and capital can be used to produce guns or butter along the frontier curve in figure 1·2. Where does the society choose to end up? Southeastward in the diagram, with much of civilian goods? Or northwestward, with much of defense goods? Economics is a quantitative subject: choice is not a qualitative matter of either-or, but rather of how many of each good and just where we draw the line of final decision.

The production-possibility frontier provides a rigorous definition of scarcity:

"Economic scarcity" refers to the basic fact of life that there exists only a finite amount of human and nonhuman resources, which the best technical knowledge is capable of using to produce only a limited maximum amount of each and every good, as shown by the p-p frontier. And thus far, nowhere on the globe is the supply of goods so plentiful or the tastes so limited that the average family can have more than enough of everything it might fancy.

The production-possibility schedule can also help make clear the three basic problems of economic life: WHAT, HOW, and FOR WHOM.

WHAT goods are produced and consumed can be depicted by the point chosen on the p-p frontier.

HOW goods are to be produced involves an efficient choice of methods and proper assignment of different amounts and kinds of limited resources to the various industries. What would happen if the men well fitted for machine-tooling of guns ended up on farms while at the same time land well fitted for butter ended up being used for gun manufacture? We would be inside the p-p frontier, not on it. Or what if government regulations made the land most suitable for corn be used for wheat production, and the land most suitable for wheat be used for corn? We would end up with less of both corn and wheat, inside the production-possibility frontier on a diagram whose axes were labeled corn and wheat. Being inside the frontier is, as we have already seen, a crime of economic inefficiency; but it need not involve any engineering inefficiency, since, on the wrongly allocated land, production might still be following the latest methods known to science.

FOR WHOM goods are to be produced cannot be discerned from the p-p frontier alone. Sometimes, though, you can make a guess from it: if you find a society on its p-p frontier with many yachts and few compact cars, you are justified in suspecting that it enjoys considerable inequality of income and wealth among persons.

Pictures in an Exhibition

Figures 1·3 to 1·7 are meant to be self-explanatory. They show that the production-possibility frontier can illustrate many familiar, but basic, economic processes. It is necessary here only to comprehend the common-sense ideas involved.

Figure 1·3 illustrates how a society consumes much food when it is poor but shifts toward comforts and luxuries as it develops.

Figure 1·4 illustrates how the electorate must choose between private goods bought at a price and public goods paid for largely by taxes.
Figure 1-3. Shift in Production-Possibility Frontier with Development

(a) Poor Nation

(b) Developed Nation

Note: Economic development and progress shifts the production-possibility frontier outward. Before development (a), the nation is so poor it must devote almost all its resources to food, enjoying few comforts. After development (b), it goes from A to B, expanding its food consumption very little compared with its increased consumption of nonnecessities. (Note that it can now have more than before of all goods if it so wishes.)

Figure 1-5 illustrates how an economy chooses between (a) current consumption goods and (b) capital goods (machines, etc.) that make possible more of both goods (consumption and capital) in the future.

Figure 1-6 shows how Economy B, blessed by scientific and engineering discoveries, might surpass A, which was showing more thrift and investing for the future, but with less progressive technology.

Finally, figure 1-7 will prepare us for the next topic, the law of diminishing returns.

The Law of Diminishing Returns

Figure 1-7 will alert us to a famous technological economic relationship, the so-called law of diminishing returns. This law states the relation—not between two goods (such as guns and butter)—but between an input of production (such as labor) and the output that it helps produce (such as butter or, in the classic examples, corn).

Here is an example to illustrate the law of diminishing returns. We make the following controlled experiment: Given a fixed amount of land, say, 100 acres, we shall first add no labor at all. We note that with zero labor input there is no corn output. So, in table 1-2, we record zero product when labor is zero.

Now we make a second related experiment. We add 1 extra unit of labor to the same fixed amount of land. How much output do we now get? Pure reason cannot tell us; we must look to the facts of the experiment. When we do, let’s say we observe that we now have produced positive output of corn.

Figure 1-4. The Ratio of Public to Private Spending in Two Economies

(a) Frontier Society

(b) Interdependent Society

Note: With prosperity comes greater emphasis on public rather than private goods. The first economy (a) is poor and dispersed, as in Daniel Boone’s frontier days: the proportion of resources going to the government public sector is low. The second economy (b) is more prosperous and chooses to spend more of its higher income on governmental services (roads, defense, research, education). In dense urban life, it has no choice but to spend on traffic lights, police, antipollution programs, and city planning.
equal to 2,000 units (bushels or whatever units you choose for corn). We now summarize the result of this second experiment: adding 1 extra unit of labor to 100 of fixed land gives us extra output of 2,000 units.

To observe the law of diminishing returns, we must make a third controlled experiment. We still hold

Figure 1-5. The Relation between Capital Formation, Current Consumption, and Future Consumption

(a) Before Thrift

(b) After Thrift

Note: Capital formation for future consumption forces sacrifice of current consumption. In (a) three countries start out even. Country 1 does no saving for the future at \( A_1 \) (merely replacing machines). Country 2 abstains modestly from consumption at \( A_2 \). Country 3, by private sacrifice or vote, is at \( A_3 \), investing much and sacrificing much of current consumption. In (b) in the next years, Country 3 has forged ahead of 2, which has moved ahead of 1. Possessing more machines for labor to work with, Country 3 now has more of both goods.

More specifically, the law of diminishing returns refers to the diminishing amount of extra output that we get when we successively add equal extra units of a varying input to a fixed amount of some other input. (Note the emphasized words.)

Figure 1-6. Economic Progress through Investment and Invention

(a) High-investment Nation (A)

(b) High-invention Nation (B)

Note: Technical invention can more than match productivity gain from thrift alone. In (a) Country A is thrifty and advances by accumulating capital goods. In (b) Country B advances even more from 1975 to 1985 because it spends more on science and technical research. From 1985 to 1995 it grows faster still by using both methods: technical progress and much capital formation.

land fixed. Once more we vary the labor input and make sure to add again exactly the same extra unit of labor as before; i.e., we now go from 1 unit of labor to 2 units of labor to match our earlier going from 0 labor to 1 labor. We breathlessly await the outcome of the experiment in terms of extra corn produced.

Shall we now have a total of 4,000 units of corn, which would, as before, represent exactly 2,000 extra units of output produced by the extra unit of the varying labor? Or shall we find diminishing returns, with the new extra unit of input adding less than the 2,000 extra units of output which was previously added?

If the law of diminishing returns does in fact hold, our experiment can have but one result. The second extra labor unit will add less extra output than did the
Figure 1-7. Illustration of Law of Diminishing Returns When Land Is Held Constant and Population Doubles

(a) Balanced Growth

(b) Fixity of Land

Note: Diminishing returns mean outputs on fixed land cannot keep pace with population.
(a) We begin initially at the lower curve A on the left. But population now doubles and is able to spread over double the previous land, leaving each county and state in the same land-labor balance as before. Hence the new curve B depicts exactly twice the scale of food and clothing production. Finally, let labor and land both rise again by the same extra amount as between A and B. We end up at C, gaining fully as much of extra products as we did before from the same balanced additions of labor and land. Note that the arrows 1, 2, and 3 on each axis show no diminishing length.

(b) We begin at the same lower curve A, repeated on the right. Now land is held constant while population doubles. Each laborer has less land to work with than under balanced growth. Hence the B' curve is below the B curve of the diagram on the left. Finally, add another equal increment of labor, still holding land constant. The new extra product is even lower, as shown by the diminishing lengths of the arrows 1, 2, and 3, depicting extra outputs. We'll soon understand the reasons for all this, after the law of diminishing returns is mastered.

Note: Diminishing returns means outputs on fixed land cannot keep pace with population.
(a) We begin initially at the lower curve A on the left. But population now doubles and is able to spread over double the previous land, leaving each county and state in the same land-labor balance as before. Hence the new curve B depicts exactly twice the scale of food and clothing production. Finally, let labor and land both rise again by the same extra amount as between A and B. We end up at C, gaining fully as much of extra products as we did before from the same balanced additions of labor and land. Note that the arrows 1, 2, and 3 on each axis show no diminishing length.

(b) We begin at the same lower curve A, repeated on the right. Now land is held constant while population doubles. Each laborer has less land to work with than under balanced growth. Hence the B' curve is below the B curve of the diagram on the left. Finally, add another equal increment of labor, still holding land constant. The new extra product is even lower, as shown by the diminishing lengths of the arrows 1, 2, and 3, depicting extra outputs. We'll soon understand the reasons for all this, after the law of diminishing returns is mastered.

Note: Diminishing returns means outputs on fixed land cannot keep pace with population.
(a) We begin initially at the lower curve A on the left. But population now doubles and is able to spread over double the previous land, leaving each county and state in the same land-labor balance as before. Hence the new curve B depicts exactly twice the scale of food and clothing production. Finally, let labor and land both rise again by the same extra amount as between A and B. We end up at C, gaining fully as much of extra products as we did before from the same balanced additions of labor and land. Note that the arrows 1, 2, and 3 on each axis show no diminishing length.

(b) We begin at the same lower curve A, repeated on the right. Now land is held constant while population doubles. Each laborer has less land to work with than under balanced growth. Hence the B' curve is below the B curve of the diagram on the left. Finally, add another equal increment of labor, still holding land constant. The new extra product is even lower, as shown by the diminishing lengths of the arrows 1, 2, and 3, depicting extra outputs. We'll soon understand the reasons for all this, after the law of diminishing returns is mastered.

The law of diminishing returns is an important, often observed economic and technical regularity. But it is not universally valid. Often it will hold only after you have added a considerable number of equal doses of the varying factor. Beyond that point, we say the law of diminishing returns has set in. (Before such a point, the varying factors might be yielding increasing extra returns, since until then we might find that adding extra varying inputs to a fixed input leads to increasing rather than diminishing extra outputs; but, ultimately, decreasing returns can be expected to prevail.)

Why is the law of diminishing returns plausible? Frequently we feel that by adding land and labor together—no input being fixed and all being varied in the same balanced proportion so that the whole scale of operations is getting larger—then output should also increase proportionately and extra outputs need not diminish. For why should the extra outputs diminish if each of the inputs always has as much of the other inputs to work with?

In short, balanced scale changes may often be expected to leave inputs and outputs in the same ratios [see figure 1-7 (a)].

By contrast, when we do hold one input or group of inputs constant and vary the remaining inputs, we see that the varying inputs have less and less of the fixed inputs to work with. Consequently, we are not too surprised that such extra varying inputs begin to add less and less extra product.

In effect, the fixed factor of production (land) is decreasing in proportion to the variable input (labor). As we crowd the land more and more, we may still get some extra corn by intensive cultivation of the soil; but the amount of extra corn per unit of extra labor will become less and less. We shall see that the real wage paid to workers depends upon the extra output a last man adds for his employer. Diminishing returns reveals that living standards in crowded China or India are low because of this basic technical truth.
Table 1-2. The Law of Diminishing Returns

<table>
<thead>
<tr>
<th>Man-years of labor</th>
<th>Total product (bushels)</th>
<th>Extra output added by additional unit of labor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2,000</td>
</tr>
<tr>
<td>1</td>
<td>2,000</td>
<td>1,000</td>
</tr>
<tr>
<td>2</td>
<td>3,000</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>3,500</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3,800</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>3,900</td>
<td></td>
</tr>
</tbody>
</table>

Note: Diminishing returns is a fundamental law of economics and technology. The law of diminishing returns refers to successively lower extra outputs (e.g., of corn) gained from adding equal increments of a variable input (e.g., labor) to a constant amount of a fixed input (e.g., 100 acres of land). (Pencil in the extra output of the fourth laborer.)

Increasing returns to scale, or so-called economies of mass production, are often associated with one of the following advances: (1) the use of nonhuman and nonanimal power sources (water and wind power, steam, electricity, turbines and internal-combustion engines, internal nuclear energy); (2) the use of automatic self-adjusting mechanisms (lathes, jigs, servomechanisms); (3) the use of standardized, interchangeable parts; (4) the breakdown of complex processes into simple repetitive operations; (5) the specialization of function and division of labor; and many other technological factors as well. The auto-production assembly line and historical development of textile spinning and weaving exemplify these diverse factors.

Upon thought, it will be evident that each of these economies or savings comes into full play only when a large enough number of units is being produced to make it worth while to set up a fairly elaborate productive organization. If only a few guns are to be produced they might just as well be produced by hand; but if resources are available to produce many thousands, it will pay to make certain elaborate initial preparations that need not be repeated when still more units are to be produced. In such cases, where mere scale matters much, the tendency for land fixity to force diminishing extra returns to labor could be thwarted for a long time by an increase in the total labor scale involved.

Economies of scale are very important in explaining why so many of the goods we buy are produced by large companies, as Karl Marx emphasized a century ago. We shall see that they are important in helping explain "the division of labor" and pattern of "specialization." They raise questions to which we shall return again and again in later chapters—as, for example, monopoly.

Economies of Scale and Mass Production: An Important Counterforce

Before leaving this section we should digress to take note of a phenomenon that is different from our controlled variation of one thing at a time.

Suppose we merely increase "scale of operations," i.e., increase all the factors at the same time in the same degree. In many industrial processes, when you double all inputs, you may find that your output is more than doubled: this phenomenon is called "increasing returns to scale."

Our previous law of diminishing returns always refers to cases where some factors were varied while some remained fixed. Hence, this case of increasing returns to scale is not a direct refutation of the law of diminishing returns.

Law of Increasing (Relative) Costs

We are now in a position to indicate why the production-possibility frontier has been drawn as a bowed-out (concave from below) curve in all our diagrams. If the p-p frontier were a straight line, the relative costs of getting some extra guns in terms of sacrificed butter would always be the same. Economists call this a case of "constant (relative) costs." But actually it is more common in life to meet the law of increasing (relative) costs.

Definition: The "law of increasing (relative) costs" prevails when in order to get equal extra amounts of one good, society must sacrifice ever-increasing amounts of the other good. A bowed-out or concave curvature of the production-possibility frontier depicts the law of increasing (relative) costs.
costs—as shown by the fact that when we want more farm goods, their (relative) cost rises in terms of sacrificed manufactures.

Why is this reasonable? We’ll see that the law of increasing (relative) costs is related to, but definitely not the same thing as, the law of diminishing returns. We’ll see that, along with the law of diminishing returns, economists must be able to assume that guns and butter use the factors of production, such as labor and land, in different proportions or intensities if we are to deduce this law of (increasing) relative costs.

To derive the law of increasing (relative) costs, let us use an oversimplified example. In figure 1-8, suppose that manufactures (guns) require labor alone and negligible land. But suppose that agricultural food (butter) requires, along with labor, fertile land that has grown scarce enough to have become private property. Now we have the two ingredients needed to explain the law of increasing (relative) costs shown in the p-p frontier: The two industries do use land and labor in different proportions; so transferring varying amounts of labor onto fixed agricultural land will turn out to bring into play the law of diminishing returns. Let us see exactly how.

Begin by using all labor to produce manufactures, at A in figure 1-8. Now sacrifice equal amounts of manufactures to get more agriculture, going to B and C. How is this transformation accomplished? By transferring equal amounts of labor away from Manufactures (since no land is used there at all). But note that these equal amounts of transferred labor are now applied to a fixed total of agricultural land. As equal amounts of varying labor are applied to a fixed land factor, each has fewer and fewer acres to work with and hence adds less and less of extra agriculture product. We see then why each new unit of agriculture is procured at higher and higher costs in terms of sacrificed manufactures.

**Efficiency**

The production-possibility frontier allows us to characterize an efficient allocation of resources as a point on the production-possibility frontier. Each point on the production-possibility frontier is a point of "technical efficiency" or maximum production in the sense of representing a maximum possible combination of the products, given the supply of resources and techniques of production. At such a point, it is impossible to produce more of one commodity without producing less of another commodity. An ideally operating price system should allocate resources so that technical efficiency is achieved.

To be distinguished from technical efficiency is "economic efficiency." This is a situation in which resources are allocated to satisfy not only the technological relations of maximum production but also the exchange relations of optimal production. Economic efficiency or the optimality of production is also designated
by some specific efficiency point on the production-possibility frontier, but at this point the production and consumption of commodities is such that it would be impossible to engage in any voluntary exchange that would make any one individual "better off" without making another individual "worse off" through a reduction of real income. Beyond technical efficiency—that is, "production-maximizing"—is economic efficiency—that is, "production-optimizing." These properties of efficiency are elaborated in detail in chapter 3.

Such a situation in which it is impossible to increase the welfare of any individual without decreasing the welfare of another individual is denoted by the economist as Pareto optimum. The precise requirements for a Pareto optimum will be stated more fully in chapter 3. As we shall see, an ideally operating price system should allocate resources so that not only is technical efficiency achieved, but also economic efficiency. A basic theorem is that under conditions of perfect competition, the working of the market-price mechanism will lead to an economically efficient outcome, a Pareto optimum (see chapter 3).

A Competitive Price System

The most significant structural and operational characteristics of a price system are: (1) division of labor and specialization of productive activities; (2) exchange transactions on markets, both product (output) markets and factor (input) markets; (3) the exercise of choice among alternatives by economic agents (households, business firms, government); and (4) adjustment to environmental change. In considering the third characteristic, the economist normally postulates rational maximizing behavior by economic agents—that is, a consumer maximizes satisfaction or preferences for a given expenditure, and firms maximize profit for a given expenditure. Considering the fourth characteristic, the economist emphasizes the environmental change of prices as leading to a change in inputs in production, a change in output, a change in demand, and a change in income distribution.

These characteristics of a price system are reflected in the operation of the forces of supply and demand. Demand and supply will determine the price in any particular market, but as the forces of demand and supply change so too does the system of prices—that is, relative prices among different markets.

The following selection (1.2) provides a synoptic view of the workings of a competitive price system.

1.2 A Picture of Prices and Markets

The Price System

Just how does the unconscious automatic price mechanism operate? The bare outlines of a competitive profit-and-loss system are simple to describe.

Everything has a price—each commodity and each service. Even the different kinds of human labor have prices, namely, wage rates. Everybody receives money for what he sells, and uses this money to buy what he wishes.

If more is wanted of any good—say, shoes—a flood of new orders will be given for it. This will cause its price to rise and more to be produced.

But what if more of a commodity—such as tea—becomes available than people want to buy at the last-quoted market price? Then its price will be marked down by competition. At the lower price people will drink more tea, and producers will no longer produce quite so much. Thus, equilibrium of supply and demand will be restored.

What is true of the markets for consumers' goods is also true of markets for factors of production such as labor, land, and capital inputs. If welders rather than glassblowers are needed, job opportunities will be more favorable in the welding field. The price of welders, their hourly wage, will tend to rise, while that of glassblowers will tend to fall. Other things being equal, this will cause a shift into the desired occupation. Likewise, an acre of land will go into sugar cultivation if sugar producers bid the most for its use. In the same way, machine-tool production will be determined by supply and demand.

THE GENERAL-EQUILIBRIUM SYSTEM

In other words, we have a vast system of trial and error, of successive approximation to an equilibrium system of prices and production. We shall see later that the matching of supply and demand and of prices and costs helps solve our three problems simultaneously. Here are the bare outlines of competitive equilibrium.

1. WHAT things will be produced is determined by the dollar votes of consumers—not every two or four years at the polls, but every day in their decisions to purchase this item and not that. Of course, the money that they pay into business cash registers ultimately provides the payrolls, rents, and dividends that consumers receive in weekly income. Thus the circle is a complete one.

2. HOW things are produced is determined by the competition of different producers. The method that is cheapest at any one time, because of both physical efficiency and cost efficiency, will displace a more costly method.

The only way for producers to meet price competition and maximize profits is to keep costs at a minimum by adopting the most efficient methods. For example, synthetic rubber will be made from oil rather than alcohol if the price of the one is in a certain relation to the price of the other; or electric power will be generated by steam rather than nuclear power if the price of coal is below some critical level. The large, tractor-operated farm will displace the family-size farm if the methods of the former lead to lower costs of production.

International example: Bob Jones farms extensively, with much American land relative to each hour of labor; Pierre Reny farms intensively, using much labor to each hectare of French land. Who orders these sensible How decisions, which properly adjust to the fact that France is more densely populated than America? Congress? The National Assembly? The UN? Of course not. The price system is society's signaling device. Like a master who gives carrots and kicks to coax his donkey forward, the pricing system deals out profits and losses to get WHAT, HOW, and FOR WHOM decided.

3. FOR WHOM things are produced is determined by supply and demand in the markets for productive services; by wage rates, land rents, interest rates, and profits, all of which go to make up everybody's income—relative to everyone else and relative to the whole. (Of course, the character of the resulting distribution of income is highly dependent upon the initial distribution of property ownership, upon acquired or inherited abilities, educational opportunities, and presence or absence of racial and sex discriminations.)

Consumer votes do not by themselves determine WHAT goods are provided. Demand has to meet with a supply of goods; so business cost and supply decisions, along with consumer demand, do help determine WHAT. Just as a broker may help arrange a match between buyer and seller, the auctioneer in the commodity market acts as the go-between who reconciles the consumer votes and business supplies that impinge on the market. The profit seeker is society's agent to determine HOW, seeking least-factor costs for producing each good and being punished by ruthless competition if he fails to use best methods.

A Picture of Prices and Markets

To amplify this highly simplified explanation, turn to figure 1-9. It gives a bird's-eye view of the way market pricing reconciles public demand and supply with business supply and demand. Note that markets serve as the connecting device between the public and business. Fifteen minutes of poring over this diagram may be worth an hour of disconnected musing about economic pricing.

A competitive system is impersonal but not completely so. The consuming families face business enterprises on two fronts, with only prices in between. One front is the widely dispersed one, the retail market on which consumers buy thousands of small items from a score of different retail establishments: grocery, drug, and department stores; movie theaters; gasoline stations; and from electric-power companies, public post offices, landlords, railroad lines, and insurance companies.

On the other front—the market for labor and other productive services—relations are not always so peaceful. To the family breadwinner his wage is not simply another price; it is the difference between luxury and comfort, between comfort and privation. The laborer may feel inferior to the large corporation in bargaining power, and he may turn to collective bargaining through trade unions. By doing this, he may at times be helping to restore competition, while
Figure 1-9. How Market Pricing Reconciles Public Demand and Supply with Business Supply and Demand

Note: The competitive price system uses supply-demand markets to solve the basic economic problems—WHAT, HOW, and FOR WHOM. See how consumer-dollar votes of demand interact in the upper goods markets with business cost-supply decisions, thus helping determine WHAT is produced. And see how business demand for inputs or productive factors meets the public's supply of labor and other inputs in the lower factor markets to help determine wage, rent, and interest income—that is, FOR WHOM goods are produced. Business competition to buy factor inputs and sell goods most cheaply determines HOW goods are to be produced. (Warning: All parts of the diagram interact together. WHAT depends on the lower part, just as FOR WHOM depends on the upper part—carpenter wages depend on housing demand, and demand for yachts depends on oil-land royalties.)

Ethical Aspects of Income Distribution

The above portrait of competition tending toward ideal efficiency, toward being on the production-possibility frontier and not inside it, is a highly oversimplified one. But even if the system worked as perfectly as described above—which all know is not the case—many would not consider it ideal.

In the first place, goods go where there are the most votes or dollars. John D. Rockefeller's dog may receive the milk that a poor child needs to avoid rickets. Why? Because supply and demand are working badly? Quite possibly badly from ethical viewpoints, but not from the standpoint of what the market mechanism is alone geared to accomplish. Functionally, auction markets are doing what they are designed to do—putting goods in the hands of those who can pay the most, who have the most money votes. Defenders and critics of the price mechanism should recognize this fact. And when a democratic
society does not like the distribution of dollar votes under laissez-faire, it uses redistributive taxation to rectify the situation.

There is another feature of even an ideal market system. Suppose the invention of automatic machines should cause the competitive price of labor to fall greatly, thereby reducing incomes of the poor. Would all ethical observers regard that as necessarily right or ideal? Certainly not.

Should the fact that a man inherited 500 square miles of rangeland, for which oil companies offer a million dollars per year, necessarily justify so large an income?

These questions are discussed repeatedly in Congress. Whether incomes should be completely determined by a competitive struggle—the survival of the survivors—is an ethical question that goes beyond the mere mechanics of economics. In the modern mixed economy, the electorate insists on providing minimum standards when the market fails to do so. Economics teaches how interventions can be accomplished at least cost in terms of inefficiency.

Imperfections of Competition

As stated earlier, one drawback to the picture of the price system as described above is the fact that, in the real world, competition is nowhere near "perfect." Firms do not know when consumer tastes will change; therefore they may overproduce in one field and underproduce in another. By the time they are ready to learn from experience, the situation may have changed again. Also, in a competitive system many producers simply do not know the methods of other producers, and costs do not fall to a minimum. In the competitive struggle, one can sometimes succeed as much by keeping knowledge scarce as by keeping production high.

The most serious deviation from perfect competition comes from monopoly elements. These—as shown later on—may result in wrong pricing, creation of distorted patterns of demand by repetitive advertising, incorrect and wasteful resource allocation, and monopoly profits. We shall be reminded again and again how strict is the economist's definition of a "perfect competitor." The mere presence of a few rivals is not enough for perfect competition.

MONOPOLY ELEMENTS

The economic definition of "imperfect competitor" is anyone who buys or sells a good in large enough quantities to be able to affect the price of that good. To some degree that means almost all businessmen, except possibly the millions of farmers who individually produce a negligible fraction of the total crop. All economic life is a blend of competitive and monopoly elements. Imperfect (monopolistic) competition is the prevailing mode, not perfect competition. A good approximation of perfect competition may be the most society can strive for.

Of course, as we shall later see, a businessman cannot set his prices completely as he pleases and still make profits. He must take into account the prices of goods that are substitutes for his own. Even if he produces a trademarked coal with unique properties, he must reckon with prices charged for other coals, oil, gas, and insulation.

Businessmen, farmers, and workers both like and dislike competition. We all like it when it enables us to expand our market, but we label it as "chiseling," "unfair," or "ruinous" when the knife cuts the other way. The worker whose livelihood depends on how the market prices his labor may be the first to howl when competition threatens to depress wages. Farm groups, aware of what competition can do to agricultural prices, bring pressure on the state to restrict production and thereby raise prices.

In the idealized model of an efficiently acting competitive market mechanism, consumers are supposed to be well informed. They recognize low quality and avoid it; they never buy drugs that turn out to be poisonous or ineffective. Most important, their "desires" are supposed to represent genuine "wants" and "needs" and "tastes." But in actual life, as Galbraith never tires of pointing out, business firms spend much money on advertising to shape—and, some insist, distort—consumer demands. We are terrorized into buying deodorants; from childhood on, we are conditioned to desire what business wants to sell.

The sequence "consumer demand→corporate price and production" is often inverted to become "corporate advertising→consumer demand→high price and profit."

Some of the basic factors responsible for monopoly-creating bigness in business may be inherent in the economies of large-scale production. This is especially true in a dynamic world of technological change. Competition by numerous producers would simply not be efficient in many fields and could not last. Trademarks, patents, and advertising are often responsible for still other market imperfections. It would be humanly impossible, therefore, to attempt to create perfect competition by law. The problem is one of achieving reasonably effective "workable competition."

A competitive system is one way of organizing an economy, but not the only way. Admiration should not inhibit reform. Still, it is of interest that some socialists plan to continue to use a price mechanism as part of their new society. A price system is not perfect, but neither are its alternatives.
Demand and Supply

We now examine the forces of demand and supply more closely. To do this, we first focus on a competitive market for one particular good (Selection 1.3). Later we shall consider the relations among different markets in a price system.

1.3 Elements of Demand and Supply

Let us start with demand. It is commonly observed that the quantity of a good that people will buy at any one time depends on price; the higher the price charged for an article, the less the quantity of it people will be willing to buy; and, other things being equal, the lower its market price, the more units of it will be demanded.

Thus there exists at any one time a definite relation between the market price of a good (such as wheat) and the quantity demanded of that good. This relation between price and quantity bought is called the "demand schedule" or "demand curve." Table 1-3 gives an example of a hypothetical demand schedule. At any price, such as $5 per bushel, there is a definite quantity of wheat that will be demanded by all the consumers in the market—in this case 9 (million) bushels per month. At a lower price, such as $4, the quantity bought is even greater, being 10 (million) units. At lower P of $3, quantity demanded is even greater still—namely 12 (million). By lowering P enough, we could coax out sales of more than 20 (million) units. From the table we can determine the quantity demanded at any price, by comparing column (2) with column (1).

The Demand Curve

The numerical data can be given a graphic interpretation also. The vertical scale in figure 1-10 represents alternative prices of wheat, measured in dollars per bushel. The horizontal scale measures the quantity of wheat (in bushels) that will be demanded per month.

A city corner is located as soon as we know its street and avenue; a ship's position is located as soon as we know its latitude and longitude. Similarly, to plot a point on this diagram, we must have two coordinate numbers: a price and a quantity. For our first point A, corresponding to $5 and 9 million bushels, we move upward 5 units and then over to the right 9 units. A dot marks the spot A. To get the next dot, at B, we go up only 4 units and over to the right 10 units. The last dot is shown by E. Through the dots we draw a smooth curve, marked dd. This illustration of the demand schedule is called the "demand curve." Note that quantity and price are inversely related, Q going up when P goes down. The curve slopes downward, going from northwest to southeast. This important property is given a name: the law of downward-sloping demand. This law is true of practically all commodities: wheat, electric razors, cotton, Kellogg's cornflakes, and theater tickets.

The law of downward-sloping demand is defined thus: When the price of a good is raised (at the same time that all other things are held constant), less of it is demanded. Or, what is the same thing: If a greater quantity of a good is put on the market, then—other things being equal—it can be sold only at a lower price.

Reasons For the Law of Downward-Sloping Demand

This law is in accordance with common sense and has been known in at least a vague way since the beginning of recorded history. The reasons for it are not difficult to identify. When the price of wheat is sky-high, only rich men will be able to afford it;
Figure 1-10. The Demand Curve

Note: Prices are measured on the vertical axis and quantities demanded on the horizontal axis. A smooth curve passed through the points gives the demand curve. The fact that dd goes downward and to the right illustrates the very important "law of downward-sloping demand."

Source: Table 1-3.

poor will have to make do with coarse rye bread, just as they still must do in poorer lands. When the price is still high but not quite so high as before, persons of moderate means who also happen to have an especially great liking for white bread will now be coaxed into buying some wheat.

Thus a first reason for the validity of the law of downward-sloping demand comes from the fact that lowering prices brings in new buyers.

Not quite so obvious is a second, equally important, reason for the law's validity; namely, each reduction of price may coax out some extra purchases by each of the good's consumers; and—what is the same thing—a rise in price may cause any of us to buy less. Why does my quantity demanded tend to fall as price rises? For two main reasons. When the price of a good rises, I naturally try to substitute other goods for it (for example, rye for wheat or tea for coffee). Also, when a price goes up, I find myself really poorer than I was before; and I will naturally cut down on my consumption of most normal goods when I feel poorer and have less real income.

Here are further examples of when I buy more of a good as it becomes more plentiful and its price drops. When water is very expensive, I demand only enough of it to drink. Then when its price drops, I buy some to wash with. At still lower prices, I resort to still other uses; finally, when it is really very cheap, I water flowers and use it lavishly for any possible purpose. (Note once again that someone poorer than me will probably begin to use water to wash his car only at a lower price than that at which I buy water for that purpose. Since market demand is the sum of all different people's demands, what does this mean? It means that even after my quantity demanded stops expanding very much with price decreases, the total bought in the market may still expand as new uses for new people come into effect.)

To confirm your understanding of the demand concept, imagine that there is an increase in demand for wheat brought about by a boom in people's incomes or by a great rise in the market price of competing corn, or simply by a change in people's tastes in favor of wheat. Show that this shifts the whole demand curve in figure 1-10 rightward, and hence upward; pencil in such a new blue curve and label it d'd' to distinguish it from the old dd curve. Note that such an increase in demand means that more will now be bought at each price—as can be verified by carefully reading off points from the new curve and filling in a new Q column for table 1-3.

The Supply Schedule

Let us now turn from demand to supply. The demand schedule related market prices and the amounts consumers wish to buy. How is the "supply schedule" defined?

By the supply schedule, or curve, is meant the relation between market prices and the amounts of the good that producers are willing to supply.

Table 1-4 illustrates the supply schedule for wheat, and figure 1-11 plots it as a supply curve. Unlike the falling demand curve, the ss supply curve for wheat

Table 1-4. Supply Schedule for Wheat

<table>
<thead>
<tr>
<th>Possible prices (dollars per bushel)</th>
<th>Quantity sellers will supply (million bushels per month)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: The table lists, for each price, the quantity that producers will want to bring to market.
normally rises upward and to the right, from southwest to northeast.

At a higher price of wheat, farmers will take acres out of corn cultivation and put them into wheat. In addition, each farmer can now afford the cost of more fertilizer, more labor, more machinery, and can now even afford to grow extra wheat on poorer land. All this tends to increase output at the higher prices offered.

Our old friend the law of diminishing returns provides one strong reason the supply curve would slope upward. If society wants more wine, then more and more labor will have to be added to the same limited hill sites suitable for producing wine grapes. Even if this industry is too small to affect the general wage rate, each new man will—according to the law of diminishing returns—be adding less and less extra product; and hence the necessary cost to coax out additional product will have to rise. (Cost and returns are opposite sides of the same coin, as will be shown later.)

4. Although exceptions to the law of downward-sloping demand are few enough to be unimportant in practice, [there is] an interesting exception to the upward-sloping supply curve. Thus, suppose that a family farmer produces wheat and its price rises so much as to give him a much higher income. With wheat so lucrative, he is at first tempted to substitute some of his leisure time to produce more. But won't there reasonably come a time when he feels comfortably enough off at his higher income to be able to afford to take things easier, work less, and supply less Q?

<table>
<thead>
<tr>
<th>Possible prices (dollars per bushel)</th>
<th>Quantity demanded (million bushels per month)</th>
<th>Quantity supplied (million bushels per month)</th>
<th>Pressure on price</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>20</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Equilibrium price is at the intersection point where supply and demand match. Only at the equilibrium price of $3, shown in the third row, will the amount supplied just match the amount demanded.

Equilibrium of Supply and Demand

Let us now combine our analysis of demand and supply to see how competitive market price is determined. This is done in table 1-5. Thus far, we have been considering all prices as possible. We have said, "If price is so and so, Q sales will be so and so; if P is such and such, Q will be such and such; and so forth." But to which level will price actually go? And how much will then be produced and consumed? The supply schedule alone cannot tell us. Neither can the demand schedule alone.

Let us do what an auctioneer would do, i.e., proceed by trial and error. Can situation A in the table, with wheat selling for $5 per bushel, prevail for any period of time? The answer is a clear "No." At $5, the producers will be supplying 18 (million) bushels to the market every month [column (3)]. But the amount demanded by consumers will be only 9 (mil-
lion) bushels per month [column (2)]. As stocks of wheat pile up, competitive sellers will cut the price a little. Thus, as column (4) shows, price will tend to fall downward. But it will not fall indefinitely to zero.

To understand this better, let us try the point $E$ with price of only $1 per bushel. Can that price persist? Again, obviously not—for a comparison of columns (2) and (3) shows that consumption will exceed production at that price. Storehouses will begin to empty, disappointed demanders who can't get wheat will tend to bid up the too-low price. This upward pressure on $P$ is shown by column (4)'s rising arrow.

We could go on to try other prices, but by now the answer is obvious:

The equilibrium price, i.e., the only price that can last, is that at which the amount willingly supplied and amount willingly demanded are equal. Competitive equilibrium must be at the intersection point of supply and demand curves.

Only at $C$, with a price of $3, will the amount demanded by consumers, 12 (million) bushels per month, exactly equal the amount supplied by producers, 12 (million). Price is at equilibrium, just as an olive at the bottom of a cocktail glass is at equilibrium, because there is no tendency for it to rise or fall. (Of course, this stationary price may not be reached at once. There may have to be an initial period of trial and error, of oscillation around the right level, before price finally is balanced.)

Figure 1-12 shows the same equilibrium in pictorial form. The supply and demand curves, superimposed on the same diagram, cross at only one intersection point. This emphasized red point $C$ represents the equilibrium price and quantity.

At a higher price, the bar shows the excess of amount supplied over amount demanded. The arrows point downward to show the direction in which price will move because of the competition of excess sellers. At a price lower than the $3 equilibrium price, the bar shows that amount demanded exceeds amount supplied. Consequently, the eager bidding of excess buyers requires us to point the arrow indicators upward to show the pressure that they are exerting on price. Only at the point $C$ will there be a balancing of forces and a stationary maintainable price.

Such is the essence of the doctrine of supply and demand.

Effect of a Shift in Supply or Demand

Now we can put the supply-and-demand apparatus to work. Gregory King, an English writer of the seventeenth century, noticed that when the harvest was bad, food rose in price; and when it was plentiful, farmers got a lower price. Let us try to explain this common-sense fact by what happens in our diagrams.
Two Stumbling Blocks

It is well to pause here to consider two minor sources of possible confusion concerning supply and demand.

Figure 1-13. How Equilibrium Price Changes When Either Supply or Demand Curve Shifts

(a) Supply Shift

(b) Demand Shift

Note: (a) If supply shifts leftward for any reason, the equilibrium-price intersection will travel up the demand curve, giving higher price and lower quantity. (b) If demand increases, the equilibrium will travel up the supply curve.

These have puzzled students of economics in all generations. The first point deals with the important fact that in drawing up a demand schedule or curve, one always insists that other things must be equal. The second deals with the exact sense in which demand and supply are equal in equilibrium.

OTHER THINGS EQUAL

To draw up a demand schedule for wheat, we vary its price and observe what would happen to its quantity bought at any one period of time in which no other factors are allowed to change so as to cloud our experiment. Specifically, this means that, as we change the price of wheat, we must not at the same time change family income or the price of a competing product such as corn or anything else that would tend to shift the demand schedule for wheat. Why? Because, like any scientist who wants to isolate the effects of one causal factor, we must try to vary only one thing at a time. True enough, in economics we cannot perform controlled experiments in a laboratory, and we can rarely hold other things constant in making statistical observations of economic magnitudes. This limitation on our ability to experiment empirically in economics makes it all the more important to be clear in our logical thinking, so that we may hope to recognize and evaluate important tendencies—such as the effect of P on Q demanded—when other tendencies are likely to be impinging on the situation at the same time.

The case of demand shift in figure 1-13(b) can illustrate this common fallacy based upon a failure to respect the rule. Other things must be held equal in defining a demand curve. Suppose that the supply curve shifts little or not at all. But suppose the demand curve shifts up to $d'd'$ in good times when jobs are plentiful and people have the incomes to buy more wheat; and suppose in the more depressed phase of the business cycle, demand always shifts down to $d'd'$. Now take a piece of graph paper and plot what would actually be recorded in the statistics of the wheat market.

In boom times, you would record the equilibrium point shown at $E'$, and in bad times the equilibrium point $E$. Take a ruler and join the points $E$ and $E'$ in Fig. 1-13(b). The fallacy to be avoided like the plague is expressed as follows: "I have disproved the law of downward-sloping demand; for note that when $P$ was high, so too was $Q$—as shown by $E'$. And when $P$ was lowered, instead of that change increasing $Q$, it actually lowered $Q$—as shown by $E$. My straight line joining $E$ and $E'$ represents an upward-sloping, not a downward-sloping, demand curve; so I have refuted a basic economic law."

Being alerted beforehand, one detects the fallacy in this argument. For at the same time that $P$ went
up, other things were not held constant; rather, income was also raised. The tendency for a rise in P to choke off purchases was more than masked by the countertendency of rising income to raise purchases. Instead of testing our economic law by moving along the demand curve, the beginner has measured changes that result from the shift of the demand curve.

Why is this bad scientific method? Because it leads to absurd results such as this: "On the basis of my revolutionary refutation of the law of downward-sloping demand I confidently predict that, in the years when the harvest is especially big, wheat will sell for a higher rather than a lower price." Not only will such reasoning lead to absurd predictions that would lose fortunes for a speculator or a miller, but it also fails to recognize other important economic relationships—such as the fact that, when family incomes go up, demand curves for goods such as wheat tend to shift rightward.

**MEANING OF EQUILIBRIUM**

The second stumbling block is a more subtle one, less likely to arise but not so easy to dispel. It is suggested by the following: "How can you say that the equality of supply and demand determines a particular equilibrium price? For, after all, the amount one man sells is precisely what another man buys. The quantity bought must always equal the quantity sold, no matter what the price; for that matter, whether or not the market is in equilibrium, a statistician who records the Q bought and the Q sold will always find these necessarily identical, each being a different aspect of exactly the same transaction."

The answer to this must be phrased something like this: "You are quite right that measured Q bought and measured Q sold must be identical as recorded by a statistician. But the important question is this: At which P will the amount that consumers are willing to go on buying be just matched by the amount that producers are willing to go on selling? At such a price, where there is equality between the scheduled amounts that suppliers and demanders want to go on buying and selling, and only at such an equilibrium P, will there be no tendency for price to rise or fall.

At any other price, such as when price is above the intersection of supply and demand, it is a trivial fact that whatever goods change hands will show a statistical identity of measured amount bought and sold. But this measured identity does not in the least deny that suppliers are eager at so high a price to sell more than demanders will continue to buy; and that this excess of scheduled supply over scheduled demand will put downward pressure on price until it has finally reached that equilibrium level where the two curves intersect.

At that equilibrium intersection, and there alone, will everybody be happy: the auctioneer, the suppliers, the demanders—as well as the patient statistician, who always reports an identity between the measured amounts bought and sold."

**What Supply and Demand Accomplished: General Equilibrium**

Having seen how supply and demand work, let us take stock of what has been accomplished. The scarce goods of society have been rationed out among the possible users of them. Who did the rationing: a board? a committee? No. The auctioneering mechanism of competitive market price did the rationing. It was a case of "rationing by the purse."

For whom goods are destined was partially determined by who was willing to pay for them. If you had the money votes, you got the wheat. If you did not, you went without. Or if you had the money votes, but preferred not to spend them on wheat, you did without. The most important needs or desires for goods—if backed by cash—got fulfilled.

The what question was being partially answered at the same time. The rise in market price was the signal to coax out a higher supply of wheat—the signal for other scarce resources to move into the wheat-production industry from alternative uses.

Even the how question was being partially decided in the background. For with wheat prices now high, farmers could afford expensive tractors and fertilizers and could bring poorer soils into use.

Why the word "partially" in this description of how the competitive market helped solve the three problems? Because this wheat market is but one market of many. What is happening in the corn and rye markets also counts; and what is happening in the market for fertilizer, men, and tractors obviously matters much. In other words, the pricing problem is one that involves interdependent markets, not just the "partial equilibrium" of a single market.

There are, so to speak, auctioneers operating simultaneously in the many different markets—wheat, rye, corn, fertilizer, and land; labor, wool, cotton, mutton, and rayon; bonds, stocks, personal loans, and foreign exchange in the form of English pounds or German marks. Each ends up at the equilibrium intersection point of his supply and demand schedules—wheat, rye, corn, fertilizer prices, and land rent; labor wage, wool, cotton, mutton, and rayon prices; bond price and its interest yield, stock prices and dividend yield, interest charges on personal loans, an exchange rate of $2.43 per pound or 3½ marks to the dollar.

No market is an island unto itself: when wool P rises (because, say, of sheep disease abroad), it pulls up the Ps of domestic labor, fertilizer, and land needed for expanded domestic wool output; and it raises the Ps of rival goods such as cotton that some demanders
Perfection and Imperfections of Competition

Our curves of supply and demand strictly apply only to a perfectly competitive market where some kind of standardized commodity such as wheat is being auctioned by an organized exchange that registers transactions of numerous buyers and sellers.

PERFECTION OF COMPETITION AS A LIMITING POLE

Needless to say, the requirements for absolutely perfect competition are as hard to meet as the requirements for a perfectly frictionless pendulum in physics. We can approach closer and closer to perfection, but can never quite reach it. Yet this fact need not do serious damage to the usefulness of our employing the idealized concept. Actually, it matters little to the economic scientist that different grades of wheat will call for slight variations from the quoted market prices. Nor does it matter in the case of standardized cotton goods, so-called gray goods, that they are sold and bought in an informal way by many competing firms; so long as there are numerous buyers and sellers on each side, well informed about quality and about each other’s prices and having no reason to discriminate in favor of one merchant rather than another and no reason to expect that variations in their own bids and offers will have an appreciable effect upon the prevailing market price—so long as all this is true, the behavior of price and quantity can be expected to be much like that predicted by our supply and demand curves.

The various diagrams in figure 1-14 illustrate how the tools of supply and demand might be used to give a good approximate description of various economic situations other than that of a staple commodity such as wheat: there is pictured a competitive market for cotton gray goods; for a factor of production such as newly graduating electrical engineers, whose price is represented by a wage per month; for a bond or capital asset such as a corner lot of land; and finally, a foreign exchange market in which the dollar price of a French franc, a German mark, or a single unit of any other foreign currency is determined by the bids of those who need foreign currency and by the offers of those who want to sell such currencies to get American dollars.

To be sure, not all today’s markets are anywhere near to being perfectly competitive in the economist’s sense. We shall see later that elements of monopoly power or of market imperfection may enter in, and these imperfections will require us to modify the competitive model. After we have learned how to handle such cases, we shall recognize that the world is a blend of competition and imperfections—which means that the competitive analysis, properly qualified, is still an indispensable tool for interpreting reality.

Of considerable importance is the responsiveness of consumers and suppliers to a change in price. Changes in the quantity demanded or supplied are measured by the coefficient of elasticity, as explained in Selection 1.4.

1.4 Elasticity of Demand and Supply

Various goods differ in the degree to which the Q bought will respond to changes in each respective P. Wheat Q may go up much less than 1 percent for each 1 percent cut in wheat P. Henry Ford’s Q may rise far more than 1 percent for each 1 percent reduction in its P. In between is the borderline case of a good whose Q just halves when its P doubles—where the percentage changes are just in balance.

Elasticity of demand is a concept devised to distinguish these three cases. Thus, the first case of weak percentage response of wheat Q is put in the category of “inelastic demand.” The second case of great percentage response is put in the category of “elastic demand.” The borderline case is called “unitary elasticity of demand.”
Figure 1-14. Supply and Demand in Various Competitive Markets

"Gray Goods"  New Electrical Engineers  Land  Foreign Exchange

<table>
<thead>
<tr>
<th>Price (per yard)</th>
<th>Wage (per month)</th>
<th>Price (per unit)</th>
<th>Price (per franc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>( d )</td>
<td>( d )</td>
<td>( d )</td>
</tr>
<tr>
<td>( E )</td>
<td>( E )</td>
<td>( E )</td>
<td>( E )</td>
</tr>
<tr>
<td>( s )</td>
<td>( s )</td>
<td>( s )</td>
<td>( s )</td>
</tr>
</tbody>
</table>

Note: Supply-and-demand tools have many applications: to goods, acres of land, laborers, exchange rates, and so on. Economists often use straight-lined \( dd \) or \( ss \) schedules purely for simplicity. Can you interpret the four cases?

Here is how the economist goes about defining the three cases:

The crucial thing to concentrate on is the total dollar revenue that buyers pay to sellers. If consumers buy five units at \$3 each, what is total revenue? Total revenue is always, by definition, price times quantity, or the \$15 product \( P \times Q \). By arithmetic multiplication, total revenue can always be calculated for each point in a demand schedule or diagram.

Elasticity of demand is important primarily as an indicator of how total revenue changes when a fall in \( P \) induces a rise in \( Q \) along the demand curve.

Definition of elasticity of demand: This is a concept devised to indicate the degree of responsiveness of \( Q \) demanded to changes in market \( P \). It depends primarily upon percentage changes and is independent of the units used to measure \( Q \) and \( P \). Elasticity ends up qualitatively in one of three alternative categories:

1. When a cut in \( P \) raises \( Q \) so much as to increase total revenue \( P \times Q \), we speak of elastic demand—or of demand elasticity greater than unity. The percentage change in \( Q \) exceeds the percentage change in \( P \). \( E_d > 1 \).

2. When a percentage cut in \( P \) results in an exactly compensating percentage rise in \( Q \) so as to leave total revenue \( P \times Q \) exactly unchanged, we speak of unitary elasticity of demand—or of demand elasticity numerically equal to unity. \( E_d = 1 \).

3. When a percentage cut in \( P \) evokes so small a percentage increase in \( Q \) as to make total revenue \( P \times Q \) fall, we speak of inelastic demand—or of demand elasticity that is less than unity (but not less than zero). \( E_d < 1 \).

Figure 1-15 gives a graphic example of the three cases. In each case, \( P \) is halved from \( A \) to \( B \), but it would be just as much in order to have used any small percentage change in \( P \). Perhaps at a first glance, it will be easiest to begin with the borderline case of unitary elasticity of demand.

In Figure 1-15(b), the doubling of \( Q \) exactly matches the halving of \( P \), with the result that the total revenue collected remains unchanged at \$1,000. This can be shown graphically by comparing certain rectangular areas. How? Price and quantity can be easily read off the curve at any point; but how do we read off total revenue, which is their arithmetic product \( P \times Q \)? When we recall that the area of a rectangle is always equal to the product of its base times its altitude, the answer is easy:

Total revenue at any point is always as shown by the area of the rectangle which that point makes with the two axes. (Check that the shaded rectangle at \( A \) does have a base equal to \( Q \) and an altitude equal to \( P \).) Hence, if our eye watches how the area of each point's rectangle changes as we cut price and move down the demand curve, we can know in which of the three categories of elasticity such a movement happens to fall.

Clearly, in the middle diagram, the shaded revenue areas remain exactly the same because of offsetting changes in their \( Q \) bases and \( P \) altitudes; consequently, this is the case neither of elastic nor of inelastic demand, but rather is the borderline case of unitary elasticity of demand.

The reader can now verify that figure 1-15(a) does correspond to elastic demand, with total revenue going up when \( P \) is cut and elasticity hence greater than unity. And figure 1-15(c) corresponds to the op-
Figure 1-15. Elasticity of Demand

Elasticity of demand comes in three cases, depending on how total revenue moves. In cutting \( P \) from \( A \) to \( B \), we raise, leave unchanged, or lower the rectangle of total revenue, depending on whether demand is elastic, unitary elastic, or inelastic. That is, elasticity depends on percentage response of \( Q \) to each percentage change in \( P \).

Do not get bogged down in numerical details of \( E_d \) calculation. Now that you have mastered the general idea of elastic, inelastic, and unitary-elastic demand, you can proceed to numerical examples.

Always there is a slight ambiguity about percentage changes. Suppose a grocer buys bread for 15 cents and sells for 25. Is that the \( 66 \frac{2}{3} \) percent markup that comes from relating the change of 10 to the lower base 15? Or is it the 40 percent change that comes from relating 10 to the higher base 25? No one answer can be said to be right, and no one definitely wrong. Fortunately, when it comes to very small percentage changes, as from 100 to 99 or from 100 to 101, the difference between \( \frac{10}{99} \) and \( \frac{10}{101} \) becomes hardly worth talking about. For small changes, it matters little how you calculate the percentage changes; but for larger ones it may make quite a difference, and no single answer can be declared to be the right one.

What is a good rule to use? As good a rule as any is to relate the price change to neither the higher nor the lower of the two \( P \)s, but to their average. Thus, is a cut from 101 to 99 a change of \( \frac{10}{99} \) or \( \frac{10}{101} \)? By our convention, it is neither: we call it a change of \( \frac{10}{100} \), because the average of 99 and 101 is \( \frac{99 + 101}{2} = \frac{200}{2} = 100 \).

Table 1-6 is self-explanatory: it shows how to calculate \( E_d \) for three movements along a dd curve. We correspond to a curve of constant elasticity, but rather one whose slope varies in order to keep the percentage changes in the same ratio. (Mathematicians call the unitary-elastic curve a "rectangular" hyperbola and know that it plots as a straight line on double-log paper.)

5. Units will affect the slope of the demand diagram, just as the draftsman can make a curve look steep or flat in slope by changing the scale of one of his axes. So the purpose of the next section is to help you avoid confusing slope and elasticity. As figure 1-15(b)'s curve with \( E_d = 1 \) shows, it is not a straight line with constant slope that

Note: Elasticity of demand comes in three cases, depending on how total revenue moves. In cutting \( P \) from \( A \) to \( B \), we raise, leave unchanged, or lower the rectangle of total revenue, depending on whether demand is elastic, unitary elastic, or inelastic. That is, elasticity depends on percentage response of \( Q \) to each percentage change in \( P \).

posite case of inelastic demand, with total revenue falling off when \( P \) is cut and elasticity less than unity.

Numerical Measurement of Elasticity: A Digression

The general notion of elastic, inelastic, and unitary elastic as an indicator of the percentage responsiveness of quantity to price and as an indicator of how total revenue behaves is now clear. But some readers will be curious to know how these qualitative cases can be given exact numerical measurement by economists. What does it mean to say that the elasticity of demand is 1.0? 2.3? 0.5? To answer this question, we give the following definition for a coefficient of demand elasticity, \( E_d \), between two different price points on a demand curve:

Elasticity coefficient \( E_d = \) \( \frac{\text{percent that } Q \text{ has risen}}{\text{percent cut in } P} \)

Note that the movements along \( P \) and \( Q \) are in opposite directions because of the law of downward-sloping demand. Note, too, the use of percentages, which brings in the nice property that the units of a good or of money—bushels or pecks of wheat, dollars or cents or francs—do not affect elasticity.
Table 1-6. Numerical Calculation of Elasticity Coefficient

<table>
<thead>
<tr>
<th>Q</th>
<th>ΔQ</th>
<th>P</th>
<th>ΔP</th>
<th>(\frac{Q_1 - Q_2}{2})</th>
<th>(\frac{P_1 + P_2}{2})</th>
<th>(\frac{\Delta Q}{\left(Q_1 + Q_2\right)/2})</th>
<th>(\frac{-\Delta P}{\left(P_1 + P_2\right)/2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>(\frac{10}{5} \div \frac{2}{5} = 5 &gt; 1)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>15</td>
<td>3</td>
<td>(\frac{10}{15} \div \frac{2}{3} = 1)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>25</td>
<td>1</td>
<td>(\frac{10}{25} \div \frac{2}{1} = .2 &lt; 1)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Dividing percentage price cut into percentage quantity rise gives numerical elasticity. Each \(-\Delta P\) is related to the average \(P\), namely, \((P_1 + P_2)/2\); each \(Q\) rise, \(\Delta Q\), to the average \(Q\), namely, \((Q_1 + Q_2)/2\); the resulting ratio gives numerical \(E_x\), a measure expressed in percentage (dimensionless) units, not in absolute slope units.

shall be seeing that most \(dd\) curves start out elastic at high \(P\) and end up inelastic at low \(P\), passing through unitary elasticity at an intermediate position where total revenue \(P \times Q\) is at its maximum. Table 1-6 illustrates this.

Graphical Measurement of Elasticity: A Digression

Students tend to make a simple mistake: They often confuse the slope of a curve with its elasticity; they think a steep slope on \(dd\) must mean inelastic demand, and a flat slope must mean elastic demand. This is not quite true. Why not? Because slope of \(dd\) depends upon absolute change in \(P\) and \(Q\), whereas elasticity was seen to depend upon percentage changes.

The straight line \(dd\) in figure 1-16(a) illustrates the fallacy of confusing slope and elasticity. Everywhere it has the same absolute slope. But toward the top of the line, where (a) \(P\) is high and its percentage change low and (b) \(Q\) is very low and its percentage change therefore almost infinitely great, our numerical formula for \(E_x\) results in a very high elasticity.

Thus, above the midpoint \(M\) of any straight line, demand is elastic, with \(E_d > 1\); at the midpoint, demand is of unitary elasticity, with \(E_d = 1\); below the midpoint, demand is inelastic, with \(E_d < 1\).^6

When many people make the same mistake, there is usually a reason. The limiting cases of completely vertical and completely horizontal demand curves, shown in figure 1-16(b) and (c), do validly portray the limiting cases of completely inelastic and infinitely elastic demands. But do not think that the in-between cases, where most of reality falls, can have their elasticities depicted by slope alone.

Now, we go back to the mainstream of demand and supply.

Elasticity of Supply

What we did for demand, we can do also for supply. Economists introduce the concept of "elasticity of supply" to give an indication of the percentage increase in the amount of \(Q\) supplied in response to a given percentage rise in competitive \(P\). (Note that in the case of a rising supply curve, we now speak of an increase in \(P\), rather than of a decrease, as in the case of a downward-sloping demand curve.)

If the amount supplied is perfectly fixed, as in the case of perishable fish brought to today's market for sale at whatever price they will fetch, we face the limiting case of perfectly inelastic, or vertical, supply. If we have a horizontal supply curve (the "constant-cost case"), so that the slightest cut in \(P\) will cause \(Q\) to become zero and the slightest rise in \(P\) will coax out an indefinitely large supply, we are at the other extreme of infinitely elastic supply. Between such extremes, we call supply elastic or inelastic depending

---

6. Intermediate books tell how to calculate \(E_d\) at any one point on a straight line: \(E_d\) equals "the length of the line segment below the point divided by the length of line segment above it." Since \(M\) is halfway, the formula there gives \(E_d\) equals 1, unitary elasticity. At \(B\), it gives \(3/1 = 3.0\); at \(R\), \(E_d = 1/3 = .33\).

Knowing how to calculate \(E_d\) for a straight line enables you to calculate it for any point along a curved \(dd\). (1) Draw with a ruler the straight line tangent to the curve at your point (e.g., at \(B\) in figure 1-17); (2) calculate the \(E_d\) for the straight line at that point (e.g., \(E_d\) at \(B = 3/1\)); (3) identify your resulting \(E_d\) as the correct elasticity for the \(dd\) curve at your chosen point.
Figure 1-16. The Distinction between Slope and Elasticity

(a) Elasticity at a Point

(b) Perfect Inelasticity

(c) Infinite Elasticity

Note: Absolute slope and percentage elasticity are not the same thing. All points on dd's straight-line demand in (a) have same absolute slope; but above the midpoint price, demand is elastic; below it, demand is inelastic; at it, demand is unitary elastic. Only in the case of perfectly vertical or perfectly horizontal curves, as in (b) and (c), can you infer inelasticity and elasticity from slope alone.

Figure 1-17. Calculating Elasticity of Demand

Note: $\eta$ at a point can be shown to be mathematically equivalent to the following limit:

$$\lim_{\Delta P \to 0} \frac{\Delta Q}{Q} = \frac{P \Delta Q}{Q \Delta P} = \frac{P}{Q} \frac{\Delta Q}{\Delta P}$$

as $\Delta P$ goes to zero, taking $\Delta Q$ with it and making it immaterial which of the $P$s and $Q$s or their averages we use to compute percentage changes. Intermediate texts show that, when you plot $dd$ on double-log paper, it becomes correct to identify slope with elastic-

tive—because double-log paper does measure percentage changes.

upon whether the percentage rise in $Q$ is respectively greater than or less than the percentage rise in $P$ bringing it about (figure 1-18).

Supply elasticity is a useful concept but not quite so useful a concept as demand elasticity, since the latter has the major additional function of telling us what is happening to total revenue.

There is, however, an important fact that supply elasticity can help describe. A given change in price will tend to have greater and greater effects on amount supplied as we move from the momentary situation to a short-run period of time and on to the long-run period. This means:

Elasticity of supply tends to be greater in the long run, when all adjustments to the higher price have been made, than in shorter periods of time.

Let us see why.

Momentary, Short-Run, and Long-Run Equilibrium

Alfred Marshall, Cambridge's great economist at the turn of the century, helped forge these tools of supply and demand. We can review our understanding of equilibrium and at the same time advance our knowl-
Figure 1-18. Supply Elasticities

Note: A numerical coefficient of supply elasticity \( E_s \) is defined thus: \( E_s = \frac{\text{percentage change in } Q}{\text{percentage change in } P} \). Figure 1-18 shows three straight-line supply curves: at A the line going through the origin has elasticity of exactly 1.0; the steeper curve with intercept on the Q axis is inelastic, with elasticity coefficient less than 1; and the flatter curve is elastic, with elasticity coefficient greater than 1. (If—as we shall see can happen—the supply curve actually bends up backward, elasticity of supply as here defined could actually become negative.) Also, for a supply curve with curvature, one can reckon its elasticity at a point \( A' \) by drawing a straight line with a tangential ruler and seeing which curve of figure 1-18 it resembles; or, plot it on double-log paper and study its slope at \( A' \), in comparison with that of a 45° line. (Verify that supply is elastic at \( A' \).)

edge if we survey Marshall's important emphasis on the time element of the problem.

He distinguishes at least three time periods: (1) momentary equilibrium, when the supply is fixed; (2) short-run equilibrium, when firms can produce more within given plants; and finally (3) long-run equilibrium (or "normal price"), when firms can abandon old plants or build new ones and when new firms can enter the industry or old ones leave it.

Let us imagine that the demand for a perishable good, such as fish that cannot be preserved, increases from \( dd \) to \( d'd' \). With the amount of fish supplied unchanged, the stronger demand will sharply bid up the momentary price of fish. This is shown in figure 1-19(a), where the fixed supply curve \( S_mS_m \) runs up to the new demand curve \( d'd' \) to determine the new, sharply higher momentary equilibrium price shown at \( E' \). The price has had to rise so much in order to ration the limited supply of fish among the now eager demanders.7

But with so high a price prevailing in the market, skippers of the fishing boats will be motivated to hire more men and to use more nets. Even if they do not have the time to get new boats built, they will in the short run begin to bring to the market a greater supply of fish than they did at the old momentary equilibrium. Figure 1-19(b) shows the new \( S_sS_s \) short-run supply schedule, and shows that it intersects the new demand curve at \( E'' \), the point of short-run equilibrium. Note that this equilibrium price is a little lower than the momentary \( E' \) price. Why? Because of the extra supply of fish induced in the short run by more intensive use of the same number of boats.

Figure 1-19(c) shows the final long-run equilibrium, or "normal," price. The higher prices that long prevailed have coaxed out more shipbuilding and attracted more trained sailors into the industry. Where the long-run supply curve \( S_lS_l \) intersects the demand curve \( d'd' \) at \( E'' \) is the final equilibrium reached after all economic conditions (including number of ships and shipyards) have adjusted to the new level of demand.

Note that the long-run equilibrium price is not so high as the short-run equilibrium price, and not nearly so high as the momentary equilibrium price. Yet it is a little bit higher than the price that prevailed previously when demand was lower. Marshall would call this a case of "increasing cost" and would regard it as the normal one to be met in most sizable competitive industries. Why normal? Because when a large industry (which has already achieved the economies of large-scale production) expands, it must coax men, ships, nets, and other productive factors away from other industries by bidding up their prices and thus its own cost. So the long-run supply curve \( S_lS_l \) will usually be sloping gently upward as in figure 1-19(c). Only if the industry is small compared with the total of all other users of its factors will Marshall's \( S_lS_l \) curve in figure 1-19(c) be horizontal—which is called the case of "constant cost."

7. If the short-run stock of goods could be carried over into the future without perishing, the \( S_mS_m \) curve would not need to be perfectly vertical. Marshall points out that people might want to reserve some of their supply for the future at low present \( P_s \), giving the \( S_mS_m \) curve some positive slope.

Applications of "the law of supply and demand" can be noted in all types of markets—for example, for the prices of products, wages, or foreign exchange rates, as already seen in Selection 1.3. Similarly, if the government intervenes in a market by establishing a price ceiling, taxing, or subsidizing, the effects of the policy can also be analyzed with simple demand-supply diagrams. This is illustrated below in Selection 1.5.
Figure 1-19. Effect of Increase in Demand on Price in Different Time Periods

(a) Momentary Equilibrium  
(b) Short-run Equilibrium  
(c) Long-run Equilibrium

Note: Effect of increase in demand on price varies in Marshall's three time periods. We distinguish between periods in which supply elements have time to make (a) no adjustments (momentary equilibrium), (b) some adjustments of labor and variable factors (short-run equilibrium), (c) full adjustment of all factors, fixed as well as varying (long-run equilibrium). The longer the time for adjustment, the greater the elasticity of supply response and the less the rise in price.

1.5 Applications of Supply and Demand

After studying supply and demand we can apply its principles to the solution of practical problems. Many people do not enjoy learning about a subject unless they can see how it "works" in the world around them. Fortunately, supply and demand analysis lends itself to a wide variety of concrete applications. Some are illustrated here in the form of real-world models.

Price Fixing by Law

A government may interfere with the normal operation of supply and demand because it wants to establish a price that is either lower or higher than that which would rule in a unregulated market. For example, price ceilings were placed on many consumer goods during World War II to keep their prices from going "too" high; today price floors keep the hourly wages of many workers from going "too" low. What are some of the economic effects of these legally established prices?

PRICE CEILINGS CAUSE SHORTAGES

The nature of a price ceiling is illustrated by the normal supply and demand curves in figure 1-20. The equilibrium price that would be established in the market if there were no outside interference would be OP (= NP'), and the equilibrium quantity would be ON. What happens if the government regards the equilibrium price as too high, and hence establishes a ceiling price making it illegal to sell the product at

a price above OH? The result will be a shortage equal to the amount RL, since this represents the excess of quantity demanded over quantity supplied at the ceiling price.

When this situation occurs, the limited supplies of the commodity OR will be snatched up by the early buyers, leaving nothing for later customers who want the remaining RL units of the product at the ceiling price. The government, therefore, may introduce rationing as an equitable method of restricting purchases. This happened during World War II, for example, when there were shortages of such price-controlled items as sugar, butter, meat, and gasoline. The government distributed ration coupons to consumers permitting them to purchase limited quantities each week.

PRICE FLOORS CAUSE SURPLUSES

Price floors are the opposite of price ceilings; they are designed to prevent a price from falling below a specified level. Although price ceilings have typically but not exclusively been a wartime phenomenon in the United States (but not in some other countries), price floors play a continuing role in our daily lives. Two types are particularly common: agricultural price supports and minimum wage legislation.

In the general model of figure 1-21 the equilibrium price and quantity that would emerge from an unregulated market are OP and ON, respectively. But now OH represents a government-imposed price floor. At this price the quantity supplied will exceed the quantity demanded, resulting in a surplus of the amount RL.

What can be done about this surplus? In agriculture, where surpluses have been a phenomenon for decades, the government has sought to cope with the situation in three major ways: (1) restrict supply by imposing acreage allotments on farmers, thereby limiting the amount of land they can use to grow certain agricultural commodities; (2) stimulate demand by encouraging research in order to find new uses for agricultural products; and (3) buy up surpluses of certain agricultural commodities and store them for future sale or disposal.

With respect to price floors, figure 1-21 may be thought of as a model of the supply and demand for labor. The horizontal axis measures the quantity of labor in hours of labor time; the vertical axis measures the price of labor in wages per hour. The surplus is then the volume of unemployment RL occurring at the minimum wage level OH. Therefore, one way to reduce this labor surplus is to lower the hourly wage rate. What would this do to total payrolls if the demand for labor were relatively elastic? Relatively inelastic? Can you suggest other possible methods of reducing the unemployment surplus?

Effects of Specific Taxes and Subsidies

Supply and demand analysis can be helpful in solving problems involving certain kinds of commodity taxes and subsidies. Different degrees of elasticities affect in surprising ways the prices and quantities of some of the things we buy every day.

A SPECIFIC TAX

Suppose a specific tax is imposed on the sale of a commodity. That is, for each unit of a commodity sold a fixed amount of money must be paid to the government. A specific tax is thus a per-unit tax independent of the price of the product. Some of the taxes on cigarettes and gasoline are of this kind.

How does a specific tax on a product affect its market prices and quantities? Is the incidence or burden of such a tax borne by those upon whom the tax is initially imposed, or is it shifted to others? These are the practical questions which our analysis will answer.

We can proceed by examining the model in figure 1-22(a). The curves DD′ and SS′ are the market demand and supply curves before the tax is imposed. The equilibrium quantity is therefore ON; the equilibrium price is NP.

Suppose that sellers are required to pay a tax of ST per unit. The results of such a tax can be analyzed in two steps:

1. The supply curve shifts to the parallel position TT′, showing that less will be supplied at any given
Figure 1-22. Effects of Specific Taxes and Subsidies

Note: Taxes will affect the equilibrium prices and quantities of commodities, depending on the relative elasticities of demand and supply. Subsidies have the opposite effects of taxes, but their influence is also determined by the relative elasticities of demand and supply.

price. This is because the tax is an added cost to the producer at all levels of output. Hence the supply price—the price necessary to call forth a given output—will be higher by the amount of the tax. For example, before the tax consumers paid a price of NP in order to obtain the quantity ON. After the tax they must pay a price of NR in order to call forth the same quantity ON. When the seller receives NR, he will pay a tax of PR (= ST) to the government, leaving himself with a net price of NP.

2. The tax will therefore cause the equilibrium point to shift from P to H. This movement will be
associated with a decrease in quantity from ON to OL and an increase in price from NP to LH, where GH is the amount of the tax.

Are there any general principles that can tell us the extent to which prices and quantities will be altered as a result of the tax? The next three charts in figure 1-22 will help answer this question.

In figure 1-22(b), since demand is perfectly elastic, any increase in price will cause sales to drop to zero. Therefore the same price is maintained after the tax as before, but sellers compensate for the added cost of the tax by reducing their quantity. Consumers thus get fewer units of the good even though they continue to pay the same price per unit.

In figure 1-22(c) the demand curve is perfectly inelastic. Therefore the entire burden of the tax is shifted forward from sellers to buyers in the form of a higher price, with no reduction in the equilibrium quantity.

In figure 1-22(d) both price and quantity are affected as a result of a perfectly elastic supply curve: the burden of the tax is shifted entirely to buyers and the equilibrium quantity is reduced. Note how this compares with the case in figure 1-22(c), where only price is affected, and not quantity. What would have happened in figure 1-22(d) if the demand curve had been perfectly inelastic?

We can now establish two important principles:

1. The more inelastic the demand and the supply of a commodity the smaller will be the decline in output resulting from a given tax. This is illustrated in figure 1-22(e), where the letters have the same meaning as before.

2. The relative burden of a tax among buyers and sellers tends to follow the path of least resistance, being shifted in proportion to where the inelasticity is greatest.

The first principle leads to the conclusion that if we want to minimize disruptions in production, industries whose commodities are inelastic in demand and supply are better suited to commodity taxation because they suffer smaller contractions in output and hence in employment.

The second principle results in the conclusion that in most supply and demand situations (except the extreme ones involving perfect elasticity or inelasticity) the tax will be shared by both consumers and producers according to the relative elasticities of demand and supply. Thus the consumer's price will rise, but by less than the amount of the tax; the producer's net price will fall, but by less than the amount of the tax.

A subsidy is a payment a government makes to individuals or businesses so that they will continue to produce a product in larger quantities or at lower prices than they would otherwise. Federal subsidies are granted to agriculture, airlines, railroads, shipping and shipbuilding, and to certain other groups in the economy.

A specific subsidy is thus a per-unit bounty; it is the opposite of a specific tax—in fact, it can be thought of as a “negative” specific tax since the government is giving money to the seller rather than taking it away.

The effects are illustrated in figure 1-22(a). As a result of a subsidy equal to the amount US, the supply curve shifts to the right from its normal position SS' to the new position UU'. This is because the subsidy is like a reduction in cost to the producer at all levels of output. Therefore his supply price will be lower by the amount of the subsidy.

For instance, the subsidy causes the equilibrium point to shift from P to K, and hence the equilibrium price to decrease from NP to MK and the equilibrium output to increase from ON to OM. At this new and larger output, buyers will pay the price MK but sellers will receive the additional amount KF (= US) which is the amount of the subsidy per unit of output.

This analysis enables us to generalize with an important principle.

The more elastic the supply and demand curves, the greater will be the expansion in output and the less will be the reduction in price resulting from a subsidy. This can be verified by comparing charts (a) and (e) in figure 1-22.

Although the economic purposes of a subsidy are to reduce price or to increase output, the latter objective is usually the primary one when the product is to be used wholly for domestic consumption. The above principle thus leads to the conclusion that if we want to increase production through the use of a subsidy, industries whose commodities are elastic in demand and supply are better suited to subsidies because they experience larger expansions in output and hence in employment.

An Ad Valorem Tax or Subsidy

The foregoing analyses can also be applied to ad valorem taxes and subsidies. An ad valorem (at value) tax is a fixed percentage of the price or value of a commodity. Therefore, unlike a specific tax, which yields eroding revenues to the government in times of inflation, an ad valorem tax is affected only by changes in the price of the product itself, hence it tends to yield tax revenues increasing at approximately the rate of inflation. Examples of ad valorem taxes are general sales taxes, property taxes, and most import duties.

A model of an ad valorem tax is presented in figure 1-23. We can analyze its effect in four basic steps.

SUBSIDIES

A subsidy is a payment a government makes to individuals or businesses so that they will continue to produce a product in larger quantities or at lower
Figure 1-23. Effect of an Ad Valorem Tax or Subsidy

![Graph showing the effects of an ad valorem tax or subsidy]

Note: The vertical distance between the old and the new supply curve must get larger with increases in output and price in order for the amount of the ad valorem tax (or subsidy) to remain a constant percentage of the selling price.

1. The original supply and demand curves are SS' and DD'. Their intersection at P determines the equilibrium output ON and the equilibrium price NP.
2. When an ad valorem tax is imposed, sellers' costs increase as a result of the tax, thereby shifting the supply curve to TT'. The equilibrium point thus changes from P to H, signifying a decrease in the equilibrium level of output from ON to OL and an increase in the equilibrium price from NP to LH. This means that in order to call forth the output OL, consumers will pay the price LH, but sellers will turn the tax GH over to the government, leaving them with a net price of LG.

So far the effects of an ad valorem tax do not seem to vary significantly from those of a specific tax as described previously. However, there is this essential difference:

3. Although an ad valorem tax is a constant percentage of the selling price, the amount of the tax, as measured by the vertical distance between the old supply curve SS' and the new supply curve TT', becomes larger with increases in output and price. (Verify this analysis by substituting your own numbers for the letters in the diagram. Once this is done, you can analyze the effects of an ad valorem subsidy in the same way.)

4. How much total revenue does the government get from the tax? You can find out by drawing the rectangle AHGB. The area of the rectangle is the number of units sold times the amount of the tax per unit, and this equals the government's total revenue. Note that the same kind of rectangle has been drawn in figure 1-22(a).

Implications for Economic Welfare

The demand schedule reflects the consumers' valuation of different quantities of the product. The supply schedule reflects the cost of supplying different quantities of the product. And this cost is based on what the resources could produce elsewhere—that is, the opportunity cost of producing this product instead of an alternative product elsewhere. Cost is a forgone benefit. The resources used to produce this particular product must therefore earn as much as they could have earned in the alternative use. It must be emphasized that this is an opportunity cost equivalent to the value of the product in the best available alternative that the community has sacrificed by using resources for this particular project rather than in the alternative.

At the equilibrium price, the valuation placed by consumers on the last unit bought (as represented by the demand schedule) is equal to the cost of producing this last unit (as represented by the supply schedule). Short of this output, the marginal value of output to consumers would exceed the alternative use value of the resources used, and output would be less than ideal; too few resources would be utilized. Beyond this output, the valuation of consumers would be less than the costs of production, and output would be greater than the ideal; too many resources would be utilized.

The competitive result of one price for all buyers is significant for economic welfare. In contrast with differential pricing—that is, different prices for different quantities purchased—uniform pricing for all quantities will leave consumers
“better off” by an amount called consumer surplus. Consumer surplus is the difference between what the consumer would have been willing to pay for the commodity rather than go without it minus what the consumer actually does pay in a single price market. According to the demand schedule in figure 1-10 (Selection 1.3), some consumers would be willing to pay $5 per bushel for 9 million bushels. But at the uniform price in equilibrium, these buyers need pay only $3 per bushel—the same as the marginal buyers. Those who would have been willing to pay $5 are therefore realizing a consumer surplus of $2 per bushel on 9 million bushels, and those who would have been willing to pay $4 are realizing a consumer surplus of $1 per bushel on an additional 1 million bushels.

The area under the demand curve to the equilibrium price line is a measure of consumer surplus. In figure 1-24 the consumer surplus is represented by the shaded area PEN when the market price is ON. If the market price should fall to ON', the consumer surplus would increase to PE'N'. An additional amount of consumer surplus (NELN') would be realized on the original quantity OX, plus the additional consumer surplus EE'L on the quantity XX'.

An increase in consumer surplus is a benefit to the private consumer. But a government may wish to value this benefit differently, depending on its concern over the distribution of income. If the government accords the same value to benefits regardless of the recipient of those benefits, the social and private measures will coincide; but the government may wish to assign a higher value to benefits accruing to poor people than those accruing to rich people. The benefit of consumer surplus would then be revalued according to its distribution among income recipients.

It is also necessary to identify the real beneficiaries of the consumer surplus on intermediate goods that enter into the production of final goods: a gain in consumer surplus by road users, for example, may in fact be an increase in their

Figure 1-24. Measurement of Consumer Surplus
profits or in the profits of middlemen or shippers or an increase in surplus for consumers of the transported goods.8

A loss in consumer surplus through a price increase can be translated into an increase in the money income that would just compensate the consumer for the price increase and fall in real income. The compensatory variation in income would offset the loss in consumer surplus and leave the consumer's real income as high as it was before the price increase. Similarly, when a price decrease raises consumer surplus and real income, a reduction in money income equivalent to the rise in consumer surplus would leave the consumer no better off than before the price reduction.

If differential pricing is practiced, the consumer surplus would be smaller than in a uniform price market. If, for example, the amount OX is sold at a price of ON while only the amount XX' is sold at the lower price ON', the consumer surplus will be reduced by NELN' as compared with the sale of the total OX' at the uniform price of ON'.

If under conditions of imperfect competition buyers can be segmented into different groups and charged different prices, some consumer surplus can be extracted for the benefit of the seller. This is the practice of price discrimination. In the extreme case, if the seller could charge a different price for each unit bought, the seller could extract all the consumer surplus by sliding down the demand curve, charging "what the traffic will bear" and hence the highest possible price for each different unit.

The contrasting effects of a uniform price and price discrimination show the value to consumers of enjoying the opportunity to buy units of a good at a particular price. Suppose the consumer were charged an "entry ticket" to buy in a uniform price market instead of being subjected to differential pricing. The consumer would then be willing to pay an amount for the entry ticket up to the amount of consumer surplus he gains through the lower uniform pricing.

Producers may also realize a surplus. The equilibrium market price reflects the price necessary to secure production from the marginal high-cost firm in the industry. But other lower-cost firms in the industry would have been willing to supply at lower prices as indicated by different points on the supply schedule. In a competitive market, however, all firms sell at the same market price, and therefore the lower-cost firms gain a surplus. The amount of producer surplus is represented by the difference between the market price line and the supply curve; in figure 1-25 the producer surplus is MPE. If price should rise to OP', the producer surplus will increase to MP'E'. If supply is perfectly inelastic, an increase in demand merely raises producer surplus, or "pure economic rent," without inducing any increase in output (see figure 1-26).

Costs of Production

So far we have been considering the supply of an industry. But the supply schedule of an industry is determined by the cost functions of the firms composing the industry. We should now consider these costs in some detail—noting especially how the economist's meaning of "cost" differs from the financial accountant's.

8. Lyn Squire and Herman G. van der Tak, Economic Analysis of Projects (Baltimore, Md.: Johns Hopkins Universi-
1.6 What Do We Mean by “Cost”?

In 1923, a famous economist by the name of J. M. Clark wrote, “A class in economics would be a success if the students gained from it an understanding of the meaning of cost in all its many aspects.” Professor Clark was prompted to make this statement by the fact that although the general idea of cost can cover a wide variety of meanings, one meaning is common to all types of cost:

Cost is a sacrifice that must be made in order to do or to acquire something. The nature of the sacrifice—i.e., what is given up—may be tangible or intangible, objective or subjective, and may take one or more of many forms such as money, goods, leisure time, income, security, prestige, power, or pleasure.

Let us amplify this definition by describing and illustrating the notion of cost.


Outlay Costs versus Opportunity (Alternative) Costs

To most of us, the concept of cost that readily comes to mind is what we may call outlay costs. These are the moneys expended in order to carry on a particular activity. Some examples of outlay costs to a business are: wages and salaries of its employees; expenditures on plant and equipment; payments for raw materials, power, light, and transportation; disbursements for rents, advertising, and insurance; and taxes paid to the government. Such costs are also frequently called explicit costs, historical costs, or accounting costs because they are the objective and tangible expenses that an accountant records in the company’s books.

There is a more basic concept of cost which economists call opportunity cost. This may be defined as the value of the benefit that is forgone by choosing one alternative rather than another. This is an extremely important concept because the “real” cost of any activity is measured by its opportunity cost, not by its outlay cost. How do you identify opportunity costs? By making a comparison between the alternative that was chosen and the one that was rejected. Here are some examples:
1. To a student, the cost of getting a full-time college education includes not only his outlay costs on tuition and books, but also the income he forgoes by not working full time.

2. To a business firm, the cost of allocating more money for advertising includes not only its outlay costs for magazine or TV time, but also the earnings it forgoes by not putting these funds to some other use—perhaps into the purchase of new equipment or the training of more salesmen.

3. To a city, the cost of a public park includes not only its outlay costs for construction and maintenance, but also the tax income that it forgoes by not zoning the land for residential, commercial, or industrial use.

You can probably think of other examples, and it should be evident why opportunity costs are often called "alternative costs."

The concept of opportunity cost arises whenever the inputs of any activity are scarce and have alternative uses. The real cost or sacrifice is then measured by the value of the forgone alternative. This principle applies at all levels of economic activity, macro as well as micro. Thus:

For any economic organism such as a society, a business, a household, or an individual, it is incorrect to confine the cost of an activity or a decision to what the organism is doing. It is what the organism is not doing but could be doing that is the correct cost consideration.

**What about Nonmonetary Alternatives?**

This principle raises an important question: Is it not true that the alternative cost of a given action may often involve nonmonetary considerations such as riskiness, working conditions, prestige, and similar factors? The answer is yes. This helps to explain why window washers in skyscrapers earn more than dishwashers in restaurants; why college professors on the average earn less—but probably have fewer headaches—than corporation executives; why the prices of "glamour" securities in the stock market fluctuate much more widely than the prices of public utility shares; and why a man may be willing to work for a smaller return in his own business where he can be

---

**Figure 1-26. Measurement of Producer Surplus When Supply Is Perfectly Inelastic**

![Diagram showing measurement of producer surplus when supply is perfectly inelastic.](image-url)
his own boss, rather than for a higher return in someone else's.

Of course, the nonmonetary elements that help make for differences in resource allocation are often difficult to measure. But in principle the monetary returns plus or minus the various nonmonetary advantages and disadvantages determine the ways in which the owners of the factors of production put their human and material resources to use.

**Economic Cost Includes Normal Profit**

Once we recognize the existence of opportunity costs, it becomes apparent that there is a sharp distinction between costs in accounting and costs in economics. Economic costs are payments that must be made to persuade the owners of the factors of production to supply them for a particular activity. This definition emphasizes the fact that economic costs are supply prices or "bids" that buyers of resources must offer to attract the factor inputs they want.

Thus a firm buys its resources such as capital, land, and labor in the open market. Its expenditures for these resources are part of its economic costs, and these money outlays are the explicit costs that an accountant records in the company's books. But there are other types of economic costs, called implicit costs because they are the costs of self-owned or self-employed resources that are not entered in a company's books of account. For example, if an individual owns a business, including the building and its real estate, and if he manages this business himself, part of his cost includes:

1. The interest return on his investment that he is forgoing by not putting his money into an alternative investment of equal risk
2. The rental receipts that he is passing up by not renting the land and building to another firm
3. The wages (including the return for entrepreneurship) that he would earn if he could be hired to manage the same kind of business for someone else.

These implicit costs of ownership constitute what may be called normal profit—that is, the least payment the owner of an enterprise would be willing to accept for performing the entrepreneurial function, including risk taking, management, and the like.

Normal profit is thus part of a firm's total economic costs, since it is a payment which the owner must receive in order to keep him from withdrawing his capital and managerial effort and putting them into some other alternative. Further, since economic costs include both explicit costs and implicit costs, and since implicit costs include normal profit, any receipts which a firm may get over and above its economic costs represent economic or pure profit.

**Short Run and Long Run**

Any discussion of costs must include an explanation of two useful concepts—the short run and the long run. These do not refer to clock or calendar time, but to the time necessary for resources to adapt fully to new conditions—regardless of how many weeks, months, or even years this may take. At any given time a firm has available a certain capacity to produce as determined by the quantity or scale of its plant and equipment. If it experiences unexpected increases or decreases in the demand for its products, it can change its level of output by using existing plant and equipment either more or less intensively, but it cannot alter plant scale or production capacity with equal speed. Business firms do not put up new factories or discard old ones with every increase or decrease in demand, any more than colleges and universities erect new classroom buildings or abandon old ones with every rise or fall in enrollment.

This leads to an important distinction between the short run and the long run. The short run is a period in which a firm can vary its output through a more or less intensive use of its resources, but cannot vary its capacity because it has a fixed plant scale. The long run is a period long enough for a firm to enter or leave an industry, and to vary its output by varying all its factors of production, including plant scale.

**The Universality of Prices**

A price system is not restricted to any one particular form of political system. Prices are crucial to the operation of the economy—whether it be a private property, market economy, or a state-owned, centrally planned economy. In a centrally planned economy prices may not be determined by the free play of market forces of demand and supply, but the establishment of prices by an official authority is still a requisite for the functioning of the economy. Whether the decisionmakers are decentralized or centralized, prices still form an organizing
device of powerful scope. For a price is a coefficient of choice that facilitates choice among alternatives.

The universality of pricing is emphasized in Selections 1.7 to 1.11.

1.7 The Economic Organization of a P.O.W. Camp

Although a prisoner-of-war (P.O.W.) camp provides a living example of a simple economy which might be used as an alternative to the Robinson Crusoe economy beloved by the textbooks, and its simplicity renders the demonstration of certain economic hypotheses both amusing and instructive, it is suggested that the principal significance is sociological. True, there is interest in observing the growth of economic institutions and customs in a brand new society, small and simple enough to prevent detail from obscuring the basic pattern and disequilibrium from obscuring the working of the system. But the essential interest lies in the universality and the spontaneity of this economic life; it came into existence not by conscious imitation but as a response to the immediate needs and circumstances. Any similarity between prison organization and outside organization arises from similar stimuli evoking similar responses.

The following is as brief an account of the essential data as may render the narrative intelligible. The camps of which the writer had experience were Oflags and consequently the economy was not complicated by payments for work by the detaining power. They consisted normally of between 1,200 and 2,500 people housed in a number of separate but intercommunicating bungalows, one company of 200 or so to a building. Each company formed a group within the main organization and inside the company the room and the messing syndicate, a voluntary and spontaneous group who fed together, formed the constituent units.

Between individuals there was active trading in all consumer goods and in some services. Most trading was for food against cigarettes or other foodstuffs, but cigarettes rose from the status of a normal commodity to that of currency; it came into existence not by conscious imitation but as a response to the immediate needs and circumstances. Any similarity between prison organization and outside organization arises from similar stimuli evoking similar responses.

The Development and Organization of the Market

Very soon after capture people realized that it was both undesirable and unnecessary, in view of the limited size and the equality of supplies, to give away or to accept gifts of cigarettes or food. "Goodwill" developed into trading as a more equitable means of maximising individual satisfaction.

We reached a transit camp in Italy about a fortnight after capture and received one-quarter of a Red Cross food parcel each a week later. At once exchanges, already established, multiplied in volume. Starting with simple direct barter, such as a non-smoker giving a smoker friend his cigarette issue in exchange for a chocolate ration, more complex exchanges soon became an accepted custom. Stories circulated of a padre who started off round the camp with a tin of cheese and five cigarettes and returned to his bed with a complete parcel in addition to his original cheese and cigarettes; the market was not yet perfect. Within a week or two, as the volume of trade grew, rough scales of exchange values came into existence. Sikhs, who had at first exchanged tinned beef for practically any other foodstuff, began to insist on jam and margarine. It was realised that a tin of jam was worth one-half pound of margarine plus something else; that a cigarette issue was worth several chocolate issues, and a tin of diced carrots was worth practically nothing.

In this camp we did not visit other bungalows very much and prices varied from place to place; hence the germ of truth in the story of the itinerant priest. By the end of a month, when we reached our permanent camp, there was a lively trade in all commodities and their relative values were well known, and expressed not in terms of one another—one didn't quote bully in terms of sugar—but in terms of cigarettes. The cigarette became the standard of value. In the permanent camp people started by wandering through the bungalows calling their offers—"cheese for seven" (cigarettes)—and the hours after parcel issue were bedlam. The inconveniences of this system soon led to its replacement by an Exchange and Mart notice board in every bungalow, where under the headings "name," "room number," "wanted," and "offered" sales and wants were advertised. When a deal went through, it was crossed off the board. The public and semipermanent records of transactions led

to cigarette prices being well known and thus tending to equality throughout the camp, although there were always opportunities for an astute trader to make a profit from arbitrage. With this development everyone, including nonsmokers, was willing to sell for cigarettes, using them to buy at another time and place. Cigarettes became the normal currency, though, of course, barter was never extinguished.

The unity of the market and the prevalence of a single price varied directly with the general level of organisation and comfort in the camp. A transit camp was always chaotic and uncomfortable: people were overcrowded, no one knew where anyone else was living, and few took the trouble to find out. Organisation was too slender to include an Exchange and Mart board, and private advertisements were the most that appeared. Consequently a transit camp was not one market but many. The price of a tin of salmon is known to have varied by 2 cigarettes in 20 between one end of a hut and the other. Despite a high level of organisation in Italy, the market was morcelated in this manner at the first transit camp we reached after our removal to Germany in the autumn of 1943. In this camp—Stalag VIIIA at Moosburg in Bavaria—there were up to 50,000 prisoners of all nationalities. French, Russians, Italians, and Yugoslavs were free to move about within the camp; British and Americans were confined to their compounds, although a few cigarettes given to a sentry would always procure permission for one or two men to visit other compounds. The people who first visited the highly organised French trading center, with its stalls and known prices, found coffee extract—relatively cheap among the tea-drinking English—commanding a fancy price in biscuits or cigarettes, and some enterprising people made small fortunes that way. (Incidentally we found out later that much of the coffee went "over the wire" and sold for phenomenal prices at black market cafes in Munich: some of the French prisoners were said to have made substantial sums in RMs. This was one of the few occasions on which our normally closed economy came into contact with other economic worlds.)

Eventually public opinion grew hostile to these monopoly profits—not everyone could make contact with the French—and trading with them was put on a regulated basis. Each group of beds was given a quota of articles to offer and the transaction was carried out by accredited representatives from the British compound, with monopoly rights. The same method was used for trading with sentries elsewhere, as in this trade secrecy and reasonable prices had a peculiar importance, but as is ever the case with regulated companies, the interloper proved too strong.

The permanent camps in Germany saw the highest level of commercial organisation. In addition to the Exchange and Mart notice boards, a shop was organised as a public utility, controlled by representatives of the Senior British Officer, on a no profit basis. People left their surplus clothing, toilet requisites, and food there until they were sold at a fixed price in cigarettes. Only sales in cigarettes were accepted—there was no barter—and there was no haggling. For food at least there were standard prices: clothing is less homogeneous and the price was decided around a norm by the seller and the shop manager in agreement; shirts would average say 80, ranging from 60 to 120 according to quality and age. Of food, the shop carried small stocks for convenience; the capital was provided by a loan from the bulk store of Red Cross cigarettes and repaid by a small commission taken on the first transactions. Thus the cigarette attained its fullest currency status, and the market was almost completely unified.

The Cigarette Currency

Although cigarettes as currency exhibited certain peculiarities, they performed all the functions of a metallic currency as a unit of account, as a measure of value and as a store of value, and shared most of its characteristics. They were homogeneous, reasonably durable, and of convenient size for the smallest or, in packets, for the largest transactions. Incidentally, they could be clipped or sweated by rolling them between the fingers so that tobacco fell out.

Cigarettes were also subject to the working of Gresham's Law. Certain brands were more popular than others as smokes, but for currency purposes a cigarette was a cigarette. Consequently buyers used the poorer qualities and the Shop rarely saw the more popular brands: cigarettes such as Churchman's No. 1 were rarely used for trading. At one time cigarettes hand-rolled from pipe tobacco began to circulate. Pipe tobacco was issued in lieu of cigarettes by the Red Cross at a rate of twenty-five cigarettes to the ounce and this rate was standard in exchanges, but an ounce would produce thirty homemade cigarettes. Naturally, people with machine-made cigarettes broke them down and re-rolled the tobacco, and the real cigarette virtually disappeared from the market. Hand-rolled cigarettes were not homogeneous and prices could no longer be quoted in them with safety: each cigarette was examined before it was accepted and thin ones were rejected, or extra demanded as a make-weight. For a time we suffered all the inconveniences of a debased currency.

Machine-made cigarettes were always universally acceptable, both for what they would buy and for themselves. It was this intrinsic value which gave rise to their principal disadvantage as currency, a disadvantage which exists, but to a far smaller extent, in the case of metallic currency—that is, a strong demand for nonmonetary purposes. Consequently our
economy was repeatedly subject to deflation and to periods of monetary stringency. While the Red Cross issue of fifty or twenty-five cigarettes per man per week came in regularly, and while there were fair stocks held, the cigarette currency suited its purpose admirably. But when the issue was interrupted, stocks soon ran out, prices fell, trading declined in volume and became increasingly a matter of barter. This deflationary tendency was periodically offset by the sudden injection of new currency. Private cigarette parcels arrived in a trickle throughout the year, but the big numbers came in quarterly when the Red Cross received its allocation of transport. Several hundred thousand cigarettes might arrive in the space of a fortnight. Prices soared, and then began to fall, slowly at first but with increasing rapidity as stocks ran out, until the next big delivery. Most of our economic troubles could be attributed to this fundamental instability.

Price Movements

Many factors affected prices, the strongest and most noticeable being the periodical currency inflation and deflation described in the last paragraphs. The periodicity of this price cycle depended on cigarette and, to a far lesser extent, on food deliveries. At one time in the early days, before any private parcels had arrived and when there were no individual stocks, the weekly issue of cigarettes and food parcels occurred on a Monday. The nonmonetary demand for cigarettes was great, and less elastic than the demand for food: consequently prices fluctuated weekly, falling towards Sunday night and rising sharply on Monday morning. Later, when many people held reserves, the weekly issue had no such effect, being too small a proportion of the total available. Credit allowed people with no reserves to meet their nonmonetary demand over the weekend.

The general price level was affected by other factors. An influx of new prisoners, proverbially hungry, raised it. Heavy air raids in the vicinity of the camp probably increased the nonmonetary demand for cigarettes and accentuated deflation. Good and bad war news certainly had its effect, and the general waves of optimism and pessimism which swept the camp were reflected in prices. Before breakfast one morning in March of this year, a rumour of the arrival of parcels and cigarettes was circulated. Within ten minutes I sold a treacle ration for four cigarettes (hitherto offered in vain for three), and many similar deals went through. By 10 o'clock the rumour was denied, and treacle that day found no more buyers even at two cigarettes.

More interesting than changes in the general price level were changes in the price structure. Changes in the supply of a commodity, in the German ration scale or in the makeup of Red Cross parcels, would raise the price of one commodity relative to others. Tins of oatmeal, once a rare and much sought after luxury in the parcels, became a commonplace in 1943, and the price fell. In hot weather the demand for cocoa fell, and that for soap rose. A new recipe would be reflected in the price level: the discovery that raisins and sugar could be turned into an alcoholic liquor of remarkable potency reacted permanently on the dried fruit market. The invention of electric immersion heaters run off the power points made tea, a drug on the market in Italy, a certain seller in Germany.

In August, 1944, the supplies of parcels and cigarettes were both halved. Since both sides of the equation were changed in the same degree, changes in prices were not anticipated. But this was not the case: the nonmonetary demand for cigarettes was less elastic than the demand for food, and food prices fell a little. More important however were the changes in the price structure. German margarine and jam, hitherto valueless owing to adequate supplies of Canadian butter and marmalade, acquired a new value. Chocolate, popular and a certain seller, and sugar, fell. Bread rose, several standing contracts of bread for cigarettes were broken, especially when the bread ration was reduced a few weeks later.

In February, 1945, the German soldier who drove the ration wagon was found to be willing to exchange loaves of bread at the rate of one loaf for a bar of chocolate. Those in the know began selling bread and buying chocolate, by then almost unsaleable in a period of serious deflation. Bread, at about forty, fell slightly; chocolate rose from fifteen; the supply of bread was not enough for the two commodities to reach parity, but the tendency was unmistakable.

The substitution of German margarine for Canadian butter when parcels were halved naturally affected their relative values, margarine appreciating at the expense of butter. Similarly, two brands of dried milk, hitherto differing in quality and therefore in price by five cigarettes a tin, came together in price as the wider substitution of the cheaper raised its relative value.

Public opinion on the subject of trading was vocal if confused and changeable, and generalisations as to its direction are difficult and dangerous. A tiny minority held that all trading was undesirable as it engendered an unsavoury atmosphere; occasional frauds and sharp practices were cited as proof. Certain forms of trading were more generally condemned: trade with the Germans was criticised by many. Red Cross toilet articles, which were in short supply and only issued in cases of actual need, were excluded from trade by law and opinion working in unshakable harmony. At one time, when there had been several cases of malnutrition reported among the more de-
voted smokers, no trade in German rations was permitted, as the victims became an additional burden on the depleted food reserves of the hospital. But while certain activities were condemned as antisocial, trade itself was practiced, and its utility appreciated, by almost everyone in the camp.

More interesting was opinion on middlemen and prices. Taken as a whole, opinion was hostile to the middleman. His function, and his hard work in bringing buyer and seller together, were ignored; profits were not regarded as a reward for labour, but as the result of sharp practices. Despite the fact that his very existence was proof to the contrary, the middleman was held to be redundant in view of the existence of an official Shop and the Exchange and Mart. Appreciation only came his way when he was willing to advance the price of a sugar ration, or to buy goods spot and carry them against a future sale. In these cases the element of risk was obvious to all, and the convenience of the service was felt to merit some reward. Particularly unpopular was the middleman with an element of monopoly, the man who contacted the ration wagon driver, or the man who utilised his knowledge of Urdu. And middlemen as a group were blamed for reducing prices. Opinion notwithstanding, most people dealt with a middleman, whether consciously or unconsciously, at some time or another.

There was a strong feeling that everything had its "just price" in cigarettes. While the assessment of the just price, which incidentally varied between camps, was impossible of explanation, this price was nevertheless pretty closely known. It can best be defined as the price usually fetched by an article in good times when cigarettes were plentiful. The "just price" changed slowly; it was unaffected by short-term variations in supply, and while opinion might be resigned to departures from the "just price," a strong feeling of resentment persisted. A more satisfactory definition of the "just price" is impossible. Everyone knew what it was, though no one could explain why it should be so.

As soon as prices began to fall with a cigarette shortage, a clamor arose, particularly against those who held reserves and who bought at reduced prices. Sellers at cut prices were criticised and their activities referred to as black market. In every period of dearth the explosive question of "should nonsmokers receive a cigarette ration?" was discussed to profitless length. Unfortunately, it was the nonsmoker, or the light smoker with his reserves, along with the hated middleman, who weathered the storm most easily.

The popularity of the price-fixing scheme, and such success as it enjoyed, were undoubtedly the result of this body of opinion. On several occasions the fall of prices was delayed by the general support given to the recommended scale. The onset of deflation was marked by a period of sluggish trade; prices stayed up but no one bought. Then prices fell on the black market, and the volume of trade revived in that quarter. Even when the recommended scale was revised, the volume of trade in the Shop would remain low. Opinion was always overruled by the hard facts of the market.

Curious arguments were advanced to justify price fixing. The recommended prices were in some way related to the calorific values of the foods offered: hence some were overvalued and never sold at these prices. One argument ran as follows: not everyone has private cigarette parcels—thus, when prices were high and trade good in the summer of 1944, only the lucky rich could buy. This was unfair to the man with few cigarettes. When prices fell in the following winter, prices should be pegged high so that the rich, who had enjoyed life in the summer, should put many cigarettes in circulation. The fact that those who sold to the rich in the summer had also enjoyed life then, and the fact that in the winter there was always someone willing to sell at low prices were ignored. Such arguments were hotly debated each night after the approach of Allied aircraft extinguished all lights at 8 P.M. But prices moved with the supply of cigarettes, and refused to stay fixed in accordance with a theory of ethics.

1.8 The Nature and Necessity of a Price System

The reason it may be useful to insert into our considerations a few remarks on the nature of price, highly theoretical though they may seem on the one hand and trivial though they may seem on the other, is simply that recent discussion on fundamental economic reform has shown that some people take the view, not new, of course, in itself, that prices and especially prices plus profits are nothing but an incident in the life of acquisitive society, that they are an obstacle to the full use of existing productive possibilities, and that they might with advantage be done away with. Prices have been compared to tolls levied for private profit or to barriers which, again for private profit, keep the potential stream of commodities from the masses who need them. The writer believes it to be a mistake to consider such views

beneath discussion and thereby to insure their survival. Among the theoretic tools needed in order to deal with this view are some of the oldest of our science, dating back to the seventeenth century and also some of the most recent ones which have been contributed to our theoretic arsenal only in the last few years. As the problems involved are familiar ground to economists, it will be possible to confine the following remarks to a few points, in fact little more than headings which could be worked out more fully.

The writer wishes to point out one thing at the outset: in the course of progress of economic analysis during the last twenty years or so, it has happened repeatedly that views largely held by practical men or amateurs which bygone generations of economists have been in the habit of disposing of as simply foolish have, by newer methods, been shown to contain some element of truth after all, and sometimes quite a large one. In no case that the writer knows of has the reasoning itself which led to such views been rehabilitated. But whilst its errors remained what they were, newer methods of analysis have repeatedly shown by other reasoning that there was yet something to the proposition which the wrong reasoning ineffectually tried to prove. It would be easy to give instances. Our problem is among them, for though as much economic insight as can be got out of an elementary course on economics would seem to be sufficient to refute that view on prices, recent investigations on limited-competition and short-period phenomena have yielded results which will go a long way toward justifying in some measure the practical implications involved in that view.

In order to show that price is a phenomenon incident to all forms of organization of society and to economic action in general, it is sufficient to look upon it as a coefficient of economic choice. That is to say, by paying a price for any commodity, buyers show a preference for that commodity as compared with other commodities which they could also buy if they wanted to, for the same money. At the point at which they stop buying, the price will exactly measure that preference for every one of them, and this is what is meant by calling price a coefficient of choice.

Now if we take the organization of a centralized socialist state as an example of non-capitalist forms of society, it stands to reason that the central management would have nothing to go by in its decisions on the questions of the what and how of production unless it gave the comrades an opportunity to express their preferences with quantitative precision. This is equivalent to saying that the coefficient of choice of the members of such a society would have to be found out somehow, for instance, by assigning to them a certain number of claims to units of product in general and allowing them to express their preferences for the various commodities by means of those units. If then prices can be considered to be coefficients of choice, then the coefficients of choice of the comrades would be essentially prices. Moreover, in order to choose between the various possible methods of production, it would be necessary for the managers to attribute values to the means of production at their command which it would be possible to deduce from the coefficients of choice expressed by the comrades. These values would be essentially the same thing as the prices of the means of production in a capitalist organization.

The last sentence already shows that the phenomenon of price covers in fact the whole range of economic action. If a man produces whisky rather than bread from his rye, then what he does can be interpreted as bartering bread for whisky, and at the point at which he stops doing this we shall again be able to obtain a quantitative expression of his preferences and again get a coefficient of choice which in all respects is the same thing as price in a market. It is obvious that the choice between these two alternatives is not determined by technical considerations. It should be equally obvious that economic considerations of precisely the same kind enter into the choice of the method of producing either bread or whisky, and that it would be incorrect to say that the decision about the what of production is an economic matter and the rest, namely, the decision about the how of production, a technological matter. For whenever there is more than one way of producing a thing, and methods of production differ as to the relative quantities of the means of production they require, it will be necessary to take account of their relative scarcity, or to put it in another way, to consider how valuable the other products are which could also be produced by the individual units of the means of production which the producer contemplates using for a given purpose. These values of alternative production show themselves in capitalist society in the money price of the means of production and would show themselves in equivalent expressions in any other form of society. This explains why technically backward methods of production may still be the most rational ones provided the more perfect methods would require less of a plentiful factor and more of one which is less plentiful, and why the technically most perfect method of production is so often a failure in economic life. Hence rational production can never rest on exclusively technological considerations, at least not as long as all means of production are not at the command of a society in unlimited quantities. An economic dimension is, therefore, always necessary for the guidance of production, and this economic dimension at all times and under all circumstances finds expression in coefficients of choice which are fundamentally the same thing as prices in a cap-
capitalist society. Of course, this does not mean that these coefficients would be numerically the same under all circumstances and in all forms of society, but they would always be of the same nature and fulfill the same purpose from which it follows that any attempt to do without them would be devoid of sense.

1.9 Economic Responsiveness in Developing Economies

Given the appropriate incentive, the quantity and value of total output in the money sector can be increased at the expense of subsistence output and leisure. A fortiori, the supply of any particular product or service in the money sector is responsive to an increase in its price relative to other prices. The time and effort which are necessary for an increase in the supply of one product can come from the same two sources as before, subsistence activities and leisure; in addition, time, effort, and money previously devoted to the production of other goods and services in the money sector can be diverted to the production of the more profitable product. Even if there were some substance in the view that backward people have an absolutely limited demand for money income or for real income generally (and this, as we have seen, is a doubtful generalization), there can be no doubt that they prefer to get the money income, and the things they wish to buy with it, as easily as possible. A relative rise in the price of one product or service is bound to increase its supply if this is physically possible.

There are various technical difficulties in the way of detailed and specific verification of this general proposition. The student of economics is familiar with the general difficulty of empirical proof because of the impossibility of eliminating the disturbing influence of factors other than those under examination, and it is distressingly easy for economists to fall into the error of confusing historical relationships with cause-and-effect relationships. Moreover, the results of the response to an increase in price may be delayed because of obstacles to an immediate increase in the rate of supply. But these and other difficulties do not detract from the validity of the proposition established from simple general reasoning, which is confirmed by the historical evidence of the increase in the production of export products and other crops, and the extension of capacity for producing these crops by peasant producers in Africa, Asia, and elsewhere under the stimulus of attractive prices. There is a wealth of examples from which to choose for purposes of illustration.

Nevertheless, policy measures influencing particular prices are frequently advocated or implemented apparently without regard for the influence of absolute and relative prices on the direction of productive effort and on investment in productive capacity; numerous examples could be cited referring to the imposition of export duties and the price policies of statutory marketing monopolies. Sometimes it is contended that the price of a certain product has risen so high as to represent the maximum possible incentive to further production. Such statements tend to neglect the possibility of extending production in outlying areas and of shifting between different activities and crops. A price which may induce a higher (perhaps even the highest possible) rate of output from a given capacity may still be insufficient to ensure an extension of productive capacity. Exceptionally high prices may be required to induce producers to establish capacity in remote areas or to switch capacity from one use to another; and these high prices, even if only temporary, may provide the wherewithal to meet the initial cost. The very high price of rubber around 1910 and in 1925–6 was largely responsible for the planting of large areas of rubber both by estates and smallholders in remote regions. Moreover, producers may be able to buy certain desired goods (such as boats, wire fencing, furniture, watches, or an expensive education for their children) only if their incomes increase quite substantially, because the goods are expensive per piece; they may be willing to make the additional effort to acquire the income only when the prices of their products are unusually attractive. Particularly high prices may be necessary to enable and to induce producers to take large strides forward in both production and consumption. Once the producers have acquired new needs and new tastes, the level of effort put forward may remain higher than before even if the high prices for their output do not persist. Further, some of the goods bought when product prices are unusually high help in subsequent production, that is, they are capital goods. In short, high prices, even though they may not be permanent, play an important part in economic development.

It may be thought that the analysis is unrealistic in assuming that prices and relative prices serve to guide productive effort in underdeveloped countries. It may be suggested that because of their illiteracy and lack of commercial sophistication, and because of the poor communications in underdeveloped countries, peasants and others in these countries cannot take advantage of changes in the relative prices and market prospects of different crops, or of the

differences in the prices for the same product offered by different buyers or in different markets, or of the rewards in different occupations and employments.

In the early stages of development from a subsistence to an exchange economy, ignorance of available economic opportunities may indeed inhibit the most profitable production or disposal of output, particularly when the hand of custom characteristic of a subsistence economy lies heavily on the producers. But once economic development has started, the sharpened interest of producers in maximising the income from their efforts, and the activities of traders in search of business enlarge knowledge of market conditions and prospects in general, and about prices and opportunities for profit in particular. Observation of behaviour in many different parts of the underdeveloped world suggests strongly that most producers are aware of current opportunities open to them, and are also anxious to use the information that they seek out or is conveyed to them.

All producers are not equally industrious and expert in the collection and interpretation of market information, or conversant with available methods of production and the costs of different methods, or agile in response to particular changes in relative prices. These are merely some of the many differences between individuals which bear on their economic performance. But it is fair to generalize that among producers and others in underdeveloped countries there is a strong tendency to make the most of economic opportunities and possibilities within the limits imposed by the state of their technical knowledge and the availability of cooperator resources.

Thus it can be concluded that people in underdeveloped countries are generally well aware of such alternatives as are open to them as sellers or buyers (though these alternatives may be limited) and of the terms on which they are open to them, and that they take advantage of changes in these alternatives, the speed of response naturally varying considerably among individuals. There is therefore no case for measures which can be justified only on the assumption that peasants and others are incapable of choosing from among available opportunities, and so are unable to make the most of them. Yet their alleged incompetence is the ostensible excuse for measures of compulsion covering the production and marketing of farm produce in many underdeveloped countries. Moreover, even if producers were ignorant, it is difficult to see how administrators can be aware of the different money costs of each producer, let alone of his preferences and of the alternative opportunities open to him. Without this knowledge any decision about what he should produce or how he should sell his output is bound to be economically wasteful from the points of view both of the producer and of the economy.

The responsiveness to changes in relative prices is in fact clearly acknowledged in the widespread practice of governments in underdeveloped countries of protecting or otherwise subsidizing the production of agricultural commodities which they wish to encourage, and of taxing those commodities the output or consumption of which they wish to discourage. The payment of premiums for certain grades of a product has often succeeded in its purpose of inducing an increase in their supply. The statutory marketing monopolies in West Africa have in this way elicited a great increase in the supply of grades of cocoa and oil-palm produce qualifying for the premiums. 9

1.10 The Market Mechanism as an Instrument of Development

In recent times, there has been a retreat both in economic theory and in economic policy from the nineteenth-century ideal of the unfettered market as a principle of economic organization. But the economic pros and cons of this retreat have been fully debated, and the economist consequently has a great deal to say about the relative merits of the market as contrasted with other methods of economic organization, and the circumstances appropriate to each.

The subject of planning and the market in economic development is, therefore, one which falls definitely within the field of the economist. Before I go on to discuss it, I must define more precisely what I mean by it. "Planning and the market" may be interpreted in two different ways. First, it may refer to the contrast between direction of the economy by government and the policy of laissez-faire. This is not my subject, though in a wider philosophical and historical context it offers much to discuss. For example, though laissez-faire and direction are often regarded as opposites, if one looks to the history of economic development one finds that economic development is almost invariably a process in which planning and direction on the one hand and freedom of enterprise on the other play their part, and are mixed. There is almost no case in which economic development has been entirely planned or entirely unplanned. The usual pattern is one of some framework of control by government, within which

9. The payment of premiums which are unrelated to differences in commercial values leads to economic waste. This is discussed in sections XI-XIII of Peter T. Bauer and Basil S. Yamey, "The Economics of Marketing Reform," Journal of Political Economy, June 1954.

the entrepreneur provides his services—a mixture of bureaucracy and enterprise, in which bureaucracy takes care of the major risks of development and enterprise faces and overcomes the minor ones. Another relevant point that Easterbrook makes is that an economy which succeeds in finding a formula for growth tends to repeat that pattern after it has become inappropriate. For example, Britain has gone on trying to work the internationally oriented pattern of its nineteenth-century development; Russia has been very successful in developing heavy industry but has not yet solved the problem of agriculture.

The alternative interpretation takes planning, in the sense of a general direction of the economy, as an established principle, and considers the market as an alternative to other and more direct means of detailed control. Given the general framework of economic planning, there is still a choice between two alternative methods of looking after the details. One is by direct detailed planning by a central authority, the other is by leaving the working out of details as far as possible to the operation of the market. (There is a third alternative, in which the government is itself the entrepreneur and investor, which I shall consider later.)

This alternative interpretation is the one I shall be using: I shall discuss the question of the market mechanism as against detailed planning as an instrument of economic development. I should like to make it clear from the start that I am going to make a strong case for the market, as the preferable instrument of economic development, on two main grounds. The first is that the achievement of the desired results by control methods is likely to be especially difficult and inefficient in an underdeveloped economy; at this point I should like to remind you that a large part of Adam Smith’s argument for laissez-faire was the inefficiency and corruption he saw in the governments of his time. The second is that the remedies for the main fault which can be found with the use of the market mechanism, its undesirable social effects, are luxuries which underdeveloped countries cannot afford to indulge in if they are really serious about attaining a high rate of development. In particular, there is likely to be a conflict between rapid growth and an equitable distribution of income; and a poor country anxious to develop would probably be well advised not to worry too much about the distribution of income.

I am going to make a fairly strong case for the market, because the market figures relatively little in the literature of economic development, and the theoretical analysis which economics has developed in relation to markets is often overlooked or disregarded.

I now want to recapitulate briefly the various economic functions of the market and the price system as a method of economic organization. I shall be brief, since the argument is a familiar one.

In the first place, the market rations supplies of consumer goods among consumers; this rationing is governed by the willingness of consumers to pay, and provided the distribution of income is acceptable it is a socially efficient process. Second, the market directs the allocation of production between commodities, according to the criterion of maximum profit, which, on the same assumption, corresponds to social usefulness. Third, the market allocates the different factors of production among their various uses, according to the criterion of maximizing their incomes. Fourth, it governs the relative quantities of specific types of labor and capital equipment made available. Fifth, it distributes income between the factors of production and therefore between individuals. Thus it solves all the economic problems of allocation of scarce means between alternative ends.

These are static functions; but the market also serves in various ways to provide incentives to economic growth. Thus the availability of goods through the market stimulates the consumer to seek to increase his income; and access to the market provides an opportunity for inventors of new goods and technical improvements to profit from their exploitation. Moreover, the market serves particularly to provide an incentive to the accumulation of capital of all kinds: first to the accumulation of personal capital in the form of trained skill, since such skill earns a higher reward; and second to the accumulation of material capital, since such capital earns an income.

The argument, then, is that a properly functioning market system would tend to stimulate both economic efficiency and economic growth. And it is important to note that the market does this automatically, while it requires no big administrative apparatus, no central decisionmaking, and very little policing other than the provision of a legal system for the enforcement of contracts.

All this sounds very impressive; but it is clearly not the whole story. What, then, are the objections to the market, how serious are they, and what should be done about them in the context of economic development? I shall discuss these questions in some detail. But first I shall state briefly the central theme of my discussion. It is that in many cases the objections to the market can be overcome by reforming specific markets, so as to bring them closer to the ideal type of market; and that to overcome other objections to the market may be very expensive and may not prove to be worthwhile—in other words, the defects of the market mechanism may on balance be more tolerable than they look at first sight.

Now, what are the objections to the market? They can, I think, be classified into two main types. One type of objection is that the market does not perform...
its functions properly. The other type of objection is that the results produced by the functioning of the market are undesirable in themselves.

I begin with the first type of objection, that the market does not perform its function properly. Here it is useful to draw a distinction between two quite different sorts of cases—those in which the market operates imperfectly, and those in which a perfectly functioning market would not produce the best results.

Imperfect operation of the market in an underdeveloped country may be attributable to ignorance, in the sense of lack of familiarity with market mechanisms and of awareness of relevant information, or to the prevalence of other modes of behavior than the rational maximization of returns from effort. In the first case, the appropriate governmental policy would seem to me to be, not to assume from the market the responsibility for allocative decisions, but to disseminate the knowledge and information required to make the market work efficiently and provide the education required to use it. The second case implies a more fundamental obstacle, not only to the use of the market but also to economic development itself, and suggests that successful economic development requires a basic change in social psychology. To my mind, it raises a serious question of fact. Is it really true that people in underdeveloped countries are strangers to the idea of maximizing gains? The idea that they are is very common in the literature and policymaking of economic development; one of its manifestations is the implicit assumption that both supplies and demands are completely price-inelastic. I am very skeptical about this, partly because of Bauer’s work and partly because at least some of the actions of governments in underdeveloped areas presuppose that even the poorest producers are susceptible to price incentives. I personally do not think one is justified in assuming as a general proposition that ignorance and illiteracy necessarily imply that men are not interested in making money. If it is true, there will be serious difficulties in the way of economic development; but again, the appropriate governmental policy would seem to be to educate the people in the practice of rational economic behavior.

Even if the market functions perfectly, it will not produce the best possible results by its own criteria if there is a difference between social and private benefit or cost. This type of case may be particularly relevant to economic development; it includes the case of increasing returns to scale, and can be extended to include the possibility that technical progress or capital accumulation tend to proceed more rapidly in industry than in agriculture. But it raises an immediate question of fact—whether divergences between social and private benefit or cost are numerous and important or not. This is an important question, but one on which we do not know very much for certain. The theory of increasing returns is logically intriguing, but the influence of increasing returns still has to be disentangled from that of technical progress in historical growth. Again, it is a fact that few advanced countries are not industrial; but this by itself does not establish the wisdom of a policy of forced industrialization in an underdeveloped country. Aside from the question of fact, the existence of divergences between social and private returns does not necessarily indicate a need for the government to replace the market mechanism; instead, the operation of the market can be perfected by the use of appropriate taxes and subsidies to offset any divergences between social and private returns.

I now turn to the second type of objection to the market, the point of which is not that the market does not work in the way it should, but that the results produced are undesirable in themselves. Here, I think, there are two major objections to the market. The first is that the income distribution produced by the market is unjust and socially undesirable. The distribution of income through the market depends on the wealth and talents of different individuals, and on their individual skill in seeing a profitable opportunity of employing their money or labor. If they make a wise or lucky choice, they may obtain a much higher income. The objection is that this method of determining the distribution of income is not just. But if you attempt to intervene in the distribution of income, you immediately encounter the problem that such intervention interferes with the efficiency of the market system. If people are not allowed to enjoy the income they could obtain by their decisions, their decisions in turn will be affected, and the efficiency of the system will be impaired. There is, therefore, a conflict between economic efficiency and social justice. The extent and importance of this conflict is likely to vary according to the state of economic development. The more advanced a country is, the more likely are its citizens to have consciences about the distribution of income, and to accept the high taxation necessary to correct it without disastrously altering their behavior; and on the other hand, the higher the level of income reached, the less serious will be any slowing down of the rate of growth brought about by redistribution policies. An advanced country can afford to sacrifice some growth for the sake of social justice. But the cost of greater equality may be great to any economy at a low level of economic development that wishes to grow rapidly, particularly as it is evident that historically the great bursts of economic growth have been associated with the prospect and the result of big windfall gains; it would therefore seem unwise for a country anxious to enjoy rapid growth to insist too strongly on policies aimed at ensuring economic
equality and a just income distribution. I should add that the problem may not be in fact as serious as I have made it out to be, since in the course of time rapid growth tends in various ways to promote a more equal distribution of wealth.

I have been discussing the objection to the results of the market system on the grounds that it produces an undesirable distribution of income. A second objection of the same sort is that the free market will not produce as high a rate of growth as is desirable. I think there is a strong case for this objection, because people's actions in regard to saving and investment depend very much on their guesses about the future. Now people are likely to know their own current requirements better than the government. But the requirements of the future have to be looked at not from the individual or family point of view or that of the nation as a collection of individuals, but from the point of view of the ongoing society. The needs of society in the future, many economists agree, tend to be underprovided for by the free market.

Even if the conclusion that state action is desirable to raise the rate of growth is accepted, this conclusion nevertheless does not carry with it a number of corollaries which are often attached to it. In particular, it does not necessarily imply that the state ought to undertake development saving and investment itself. Private enterprise may be more efficient than the government in constructing and operating enterprises, so that the best policy may be to stimulate private enterprise by tax concessions, subsidies, and the provision of cheap credit. Similarly, it may be preferable to stimulate private saving by offering high interest rates, rather than by forcing savings into the hands of the state by taxation or inflation. One argument against a policy of low interest rates and forced saving is that it may in the long run contribute to the inequality of income distribution. The reason is that the poor or small savers are mainly confined to low-yielding fixed-interest investments, directly or indirectly in government debt, because these are safe and easily available, whereas the larger savers can invest their money in higher-yielding stocks and shares or directly in profitable enterprises. There is, therefore, an opportunity here for government both to stimulate saving for development and to improve the distribution of income.

There is another reason for being wary of the proposition that the state should undertake development investment itself—the danger that if the government undertakes investment itself, especially if its administrators are not too clear on their objectives, the result will be the creation of vested industrial interests inimical to further development, and resistant to technical change.

To summarize the foregoing argument from the point of view of development policy, it seems to me that much of development planning could usefully be devoted to the improvement and strengthening of the market system. This does not imply the acceptance of all the results of laissez-faire, especially with respect to the rate of growth; but there are reasons for thinking that too much emphasis on a fair or ethical distribution of income can be an obstacle to rapid growth.

The argument I have presented has been concerned mainly with one side of the case for the market. The other side concerns the costs and difficulties of controls, in terms of the manpower costs of the administration they require, and their effects in creating profit opportunities which bring windfall gains to some members of the community and create incentives to evasion which in turn require policing of the controls. I have touched on that side of the argument sufficiently frequently to make it unnecessary to elaborate on it further.

Instead, I shall comment briefly on international markets in relation to economic development, since so far I have been implicitly concerned with internal markets. Economic development planning inevitably has a strong autarkic bias, by reason both of its motivation and of the limitation of the scope of control to the national economy. Nevertheless, international trade can play an important part in stimulating and facilitating the development process. Access to foreign markets for exports can permit an economy with a limited domestic market to exploit economies of scale, and the potentiality of such exports can serve as a powerful attraction for foreign capital and enterprise. Similarly, the capacity to import provided by exports can give a developing economy immediate access to the products of advanced technology, without obliging it to go through the long and perhaps costly process of developing domestic production facilities. Economic nationalism and excessive fear of the risks of international trade, by fostering aversion to exploiting the advantages of the international market, can therefore retard economic development unnecessarily.

One further comment on the international aspects of the market and economic development seems worth making. Discussion of the international side of development has been mostly concerned with commodity trade and commercial policy. But in fact one of the most important ways in which the world market system is imperfect is with respect to the international mobility of capital and labor. The problem of international capital movements has received a fair amount of attention, labor mobility and immobility much less. The process of economic development in the past, especially in the nineteenth century, was characterized by vast movements, not only of capital but also of labor, about the world. The mass movement of labor between countries has now
been more or less shut off by the growth of nationalism. I believe it is important to recognize this restriction on international competition, and its implications for programs of economic development. It means—looking at the world economy as a whole—that the solution to the problem of maximizing world output cannot be approached directly, by bringing labor, capital technology, and natural resources together at the most efficient location; instead, the other productive factors have to be brought to the labor. To a large extent, "the economic development of underdeveloped countries" is a second-best policy, in which gifts of capital and technical training by advanced to underdeveloped countries are a compensation for the unwillingness of the former to consider the alternative way of improving the labor to resources ratio, movement of the labor to the resources. The fact that development is a second-best policy in this respect may impose severe limitations on its efficiency and rapidity.

To conclude, I have been concerned with the role of the market in economic development; and I have aimed at stressing the economic functions of the market, in automatically taking decisions about various kinds of allocations of economic resources, and the place in economic development programs of improvements in market organization and methods. I have been advocating, not a policy of laissez-faire, but recognition of the market as an administrative instrument that is relatively cheap to operate and may therefore be efficient in spite of objectionable features of its operations. The general assumption on which I have been arguing is that economic development is a process of cooperation between the state and private enterprise, and that the problem is to devise the best possible mixture.

1.11 Prices in a Socialist Economy

Ludwig von Mises' contention that a socialist economy cannot solve the problem of rational allocation of its resources is based on a confusion concerning the nature of prices. As Wicksteed has pointed out, the term "price" has two meanings. It may mean either price in the ordinary sense, that is, the exchange ratio of two commodities on a market, or it may have the generalized meaning of "terms on which alternatives are offered." Wicksteed says, " 'Price,' then, in the narrower sense of 'the money for which a material thing, a service, or a privilege can be obtained,' is simply a special case of 'price' in the wider sense of 'the terms on which alternatives are offered,' "10 It is only prices in the generalized sense that are indispensable to solving the problem of allocation of resources. The economic problem is a problem of choice between alternatives. To solve the problem three data are needed: (1) a preference scale that guides the acts of choice; (2) knowledge of the "terms on which alternatives are offered"; and (3) knowledge of the amount of resources available. Those three data being given, the problem of choice is soluble.

Now it is obvious that a socialist economy may regard the data under 1 and 3 as given, at least in as great a degree as they are given in a capitalist economy. The data under 1 may be given either by the demand schedules of the individuals or be established by the judgment of the authorities administering the economic system. The question remains whether the data under 2 are accessible to the administrators of a socialist economy. Professor Mises denies this. Nevertheless, a careful study of price theory and of the theory of production convinces us that, the data under 1 and under 3 being given, the "terms on which alternatives are offered" are determined ultimately by the technical possibilities of transformation of one commodity into another, i.e., by the production functions. The administrators of a socialist economy will have exactly the same knowledge, or lack of knowledge, of the production functions as the capitalist entrepreneurs have.

But Professor Mises seems to have confused prices in the narrower sense, i.e., the exchange ratios of commodities on a market, with prices in the wider sense of "terms on which alternatives are offered." Since, in consequence of public ownership of the means of production, there is in a socialist economy no market on which capital goods are actually exchanged, there are obviously no prices of capital goods in the sense of exchange ratios on a market. And hence, Professor Mises argues, there is no "index of alternatives available in the sphere of capital goods." But this conclusion is based on a confusion of "price" in the narrower sense with "price" in the wider sense of an index of alternatives. It is only in the latter sense that "prices" are indispensable for the allocation of resources, and on the basis of the technical possibilities of transformation of one commodity into another they are also given in a socialist economy. Professor Mises argues that private ownership of the means of production is indispensable for a rational allocation of resources. Since, according to

10. P. H. Wicksteed, The Common Sense of Political Economy, 2d ed. (London, 1933), p. 28. Similarly Schumpeter has stated that the term "exchange ratio" may be used in a wider sense to indicate the alternatives available, so that production may be regarded as an "exchange" sui generis.
him, without private ownership of the means of production no determinate index of alternatives exists (at least in the sphere of capital goods), the economic principles of choice between different alternatives are applicable only to a special institutional set-up, i.e., to a society which recognizes private ownership of the means of production. It has been maintained, indeed, by Marx and by the historical school (insofar as the latter recognized any economic laws at all) that all economic laws have only historico-relative validity. But it is most surprising to find this institutionalist view supported by a prominent member of the Austrian school, which did so much to emphasize the universal validity of the fundamental principles of economic theory.

Thus Professor Mises' denial of the possibility of economic calculation in a socialist system must be rejected. However, Professor Mises' argument has been taken up recently in a more refined form by Professor Friedrich von Hayek and Professor Lionel Robbins. They do not deny the theoretical possibility of a rational allocation of resources in a socialist economy; they only doubt the possibility of a satisfactory practical solution of the problem. Discussing the solution offered by Barone, Dickinson, and others, Professor Hayek says that "it must be admitted that this is not an impossibility in the sense that it is logically contradictory." But he denies that the problem can have a practical solution in a society without private ownership of the means of production. 11

The issue has been put very clearly by Professor Robbins. "On paper," he says, "we can conceive this problem to be solved by a series of mathematical calculations . . . But in practice this solution is quite unworkable. It would necessitate the drawing up of millions of equations on the basis of millions of statistical data based on many more millions of individual computations. By the time the equations were solved, the information on which they were based would have become obsolete and they would need to be calculated anew. The suggestion that a practical solution of the problem is possible on the basis of the Pareto equations simply indicates that those who put it forward have not grasped what these equations mean." 12

Thus Professor Hayek and Professor Robbins have given up the essential point of Professor Mises' position and retreated to a second line of defense. In principle, they admit, the problem is soluble, but it is to be doubted whether in a socialist community it can be solved by a simple method of trial and error, as it is solved in the capitalist economy. The significance of the private ownership of the means of production and of an actual market for capital goods has shifted. Theoretically prices in the generalized sense of "terms on which alternatives are offered" are admitted to be given also without an actual market. The function of the market is, according to them, a different one, namely, to provide a method of allocating resources by trial and error. And it is this function a socialist economy would be deprived of.

The position taken by Professor Hayek and by Professor Robbins is a significant step forward in the discussion of the problem. It promises a much more fruitful approach than Professor Mises' wholesale denial of the possibility of economic accounting under socialism. Whether by having taken this step they, too, will merit an honorable statue, or at least a memorial tablet, in the building of the Ministry of Socialization or of the Central Planning Board is yet to be seen. The great importance of the problem makes it quite possible.

Barone has already pointed to the fact that the equations of economic equilibrium must be solved also in a socialist society by trial and error. 13 He regarded such a solution as possible but failed to indicate how it would be achieved. However, the way in which a socialist economy would solve the problem by a method of trial and error has been indicated quite clearly by Fred M. Taylor in a paper published in 1929. 14 This paper provides in substance the answer to Professor Hayek's and Professor Robbins' argument, and it is the first contribution which really goes beyond what is contained in Barone's paper. But the great importance of the argument of Hayek and Robbins necessitates a more detailed investigation of the problem. It is, therefore, the purpose of the present essay to elucidate the way in which the allocation of resources is effected by trial and error on a competitive market and to find out whether a similar trial and error procedure is not possible in a socialist economy.

The Trial and Error Procedure in a Socialist Economy

In order to discuss the method of allocating resources in a socialist economy we have to state what kind of socialist society we have in mind. The fact of public

ownership of the means of production does not in itself determine the system of distributing consumers' goods and of allocating people to various occupations, nor the principles guiding the production of commodities. Let us now assume that freedom of choice in consumption and freedom of choice of occupation are maintained and that the preferences of consumers, as expressed by their demand prices, are the guiding criteria in production and in the allocation of resources. Later we shall pass to the study of a more centralized socialist system.

In the socialist system as described we have a genuine market (in the institutional sense of the word) for consumers' goods and for the services of labor. But there is no market for capital goods and productive resources outside of labor. The prices of capital goods and productive resources outside of labor are thus prices in the generalized sense, i.e., mere indexes of alternatives available, fixed for accounting purposes. Let us see how economic equilibrium is determined in such a system. Just as in a competitive individualist regime, the determination of equilibrium consists of two parts. (A) On the basis of given indexes of alternatives (which are market prices in the case of consumers' goods and the services of labor and accounting prices in all other cases) both the individuals participating in the economic system as consumers and as owners of the services of labor and the managers of production and of the ultimate resources outside of labor (i.e., of capital and of natural resources) make decisions according to certain principles. These managers are assumed to be public officials. (B) The prices (whether market or accounting) are determined by the condition that the quantity of each commodity demanded is equal to the quantity supplied. The conditions determining the decisions under A form the subjective, while that under B is the objective, equilibrium condition. Finally, we have also a condition C, expressing the social organization of the economic system. Since the productive resources outside of labor are public property, the incomes of the consumers are divorced from the ownership of those resources and the form of condition C (social organization) is determined by the principles of income formation adopted.

The possibility of determining condition C in different ways gives to a socialist society considerable freedom in matters of distribution of income. But the necessity of maintaining freedom in the choice of occupation limits the arbitrary use of this freedom, for there must be some connection between the income of a consumer and the services of labor performed by him. It seems, therefore, convenient to regard the income of consumers as composed of two parts: one part being the receipts for the labor services performed and the other part being a social dividend constituting the individual's share in the income derived from the capital and the natural resources owned by society. We assume that the distribution of the social dividend is based on certain principles, reserving the content of those principles for later discussion. Thus condition C is determinate and determines the incomes of the consumers in terms of prices of the services of labor and social dividend, which, in turn, may be regarded as determined by the total yield of capital and of the natural resources and by the principles adopted in distributing this yield.\textsuperscript{15}

A. Let us consider the subjective equilibrium condition in a socialist economy:

1. Freedom of choice in consumption being assumed,\textsuperscript{16} this part of the subjective equilibrium condition of a competitive market applies also to the market for consumer goods in a socialist economy. The incomes of the consumers and the prices of consumer goods being given, the demand for consumer goods is determined.

2. The decisions of the managers of production are no longer guided by the aim of maximizing profit. Instead, certain rules are imposed on them by the Central Planning Board which aim at satisfying consumers' preferences in the best way possible. These rules determine the combination of factors of production and the scale of output.

One rule must impose the choice of the combination of factors which minimizes the average cost of production. This rule leads to the factors being combined in such proportion that the marginal productivity of that amount of each factor which is worth a unit of money is the same for all factors. This rule is addressed to whoever makes decisions involving the problem of the optimum combination of factors, i.e., to managers responsible for running existing plants and to those engaged in building new plants. A second rule determines the scale of output by stating that output has to be fixed so that marginal cost

\textsuperscript{15} In formulating condition C, capital accumulation has to be taken into account. Capital accumulation may be done either "corporately" by deducting a certain part of the national income before the social dividend is distributed, or it may be left to the savings of individuals, or both methods may be combined. But "corporate" accumulation must certainly be the dominant form of capital formation in a socialist economy.

\textsuperscript{16} Of course there may be also a sector of socialized consumption, the cost of which is met by taxation. Such a sector exists also in capitalist society and comprises the provision not only of collective wants, in Cassel's sense, but also of other wants whose social importance is too great to be left to the free choice of individuals (for instance, free hospital service and free education). But this problem does not represent any theoretical difficulty and we may disregard it.
is equal to the price of the product. This rule is addressed to two kinds of persons. First of all, it is addressed to the managers of plants and thus determines the scale of output of each plant and, together with the first rule, its demand for factors of production. The first rule, to whomever addressed, and the second rule when addressed to the managers of plants perform the same function that in a competitive system is carried out by the private producer’s aiming to maximize his profit, when the prices of factors and of the product are independent of the amount of each factor used by him and of his scale of output.

The total output of an industry has yet to be determined. This is done by addressing the second rule also to the managers of a whole industry (e.g., to the directors of the National Coal Trust) as a principle to guide them in deciding whether an industry ought to be expanded (by building new plants or enlarging old ones) or contracted (by not replacing plants which are wearing out). Thus each industry has to produce exactly as much of a commodity as can be sold or “accounted for” to other industries at a price which equals the marginal cost incurred by the industry in producing this amount. The marginal cost incurred by an industry is the cost to that industry (not to a particular plant) of doing whatever is necessary to produce an additional unit of output, the optimum combination of factors being used. This may include the cost of building new plants or enlarging old ones.17

Addressed to the managers of an industry, the second rule performs the function which under free competition is carried out by the free entry of firms into an industry or their exodus from it: i.e., it determines the output of an industry.18 The second rule, however, has to be carried out irrespective of whether average cost is covered or not, even if it should involve plants or whole industries in losses.

Both rules can be put in the form of the simple request to use always the method of production (i.e., combination of factors) which minimizes average cost and to produce as much of each service or commodity as will equalize marginal cost and the price of the product, this request being addressed to whoever is responsible for the particular decision to be taken. Thus the output of each plant and industry and the total demand for factors of production by each industry are determined. To enable the managers of production to follow these rules the prices of the factors and of the products must, of course, be given. In the case of consumers’ goods and services of labor they are determined on a market; in all other cases they are fixed by the Central Planning Board. Those prices being given, the supply of products and the demand for factors are determined.

The reasons for adopting the two rules mentioned are obvious. Since prices are indexes of terms on which alternatives are offered, that method of production which will minimize average cost will also minimize the alternatives sacrificed. Thus the first rule means simply that each commodity must be produced with a minimum sacrifice of alternatives. The second rule is a necessary consequence of following consumers’ preferences. It means that the marginal significance of each preference that is satisfied has to be equal to the marginal significance of the alternative preferences the satisfaction of which is sacrificed. If the second rule were not observed, certain lower preferences would be satisfied while preferences higher up on the scale would be left unsatisfied.

3. Freedom of choice of occupation being assumed, laborers offer their services to the industry or occupation paying the highest wages. For the publicly owned capital and natural resources a price has to be fixed by the Central Planning Board with the provision that these resources can be directed only to industries which are able to “pay,” or rather to “account for,” this price. This is a consequence of following the consumers’ preferences. The prices of the services of the ultimate productive resources being

17. Since in practice such marginal cost is not a continuous function of output we have to compare the cost of each additional indivisible input with the receipts expected from the additional output thus secured. For instance, in a railway system as long as there are unused carriages the cost of putting them into use has to be compared with the additional receipts which may be obtained by doing so. When all the carriages available are used up to capacity, the cost of building and running additional carriages (and locomotives) has to be compared with the additional receipts expected to arise from such action. Finally, the question of building new tracks is decided upon the same principle. See A. P. Lerner, “Statics and Dynamics in Socialist Economics,” Economic Journal, 47 (June 1937):263–67.

18. The result, however, of following this rule coincides with the result obtained under free competition only in the case of constant returns to the industry (that is, a homogeneous production function of the first degree). In this case marginal cost incurred by the industry equals average cost. In all other cases the results diverge, for under free competition the output of an industry is such that average cost equals the price of the product, while according to our rule it is marginal cost (incurred by the industry) that ought to be equal to the price. This difference results in profits being made by the industries whose marginal cost exceeds average cost, whereas the industries in which the opposite is the case incur losses. These profits and losses correspond to the taxes and bounties proposed by Professor Pigou in order to bring about under free competition the equality of private and social marginal net product. See A. C. Pigou, The Economics of Welfare, 3d ed. (London: Macmillan, 1929), pp. 223–27.
given, their distribution between the different industries is also determined.

B. The subjective equilibrium condition can be carried out only when prices are given. This is also true of the decisions of the managers of production and of the productive resources in public ownership. Only when prices are given can the combination of factors which minimizes average cost, the output which equalizes marginal cost and the price of the product, and the best allocation of the ultimate productive resources be determined. But if there is no market (in the institutional sense of the word) for capital goods or for the ultimate productive resources outside of labor, can their prices be determined objectively? Must not the prices fixed by the Central Planning Board necessarily be quite arbitrary? If so, their arbitrary character would deprive them of any economic significance as indexes of the terms on which alternatives are offered. This is, indeed, the opinion of Professor Mises.19 And the view is shared by G. D. H. Cole, who says: "A planless economy, in which each entrepreneur takes his decisions apart from the rest, obviously confronts each entrepreneur with a broadly given structure of costs, represented by the current level of wages, rent, and interest . . .

In a planned socialist economy there can be no objective structure of costs. Costs can be imputed to any desired extent . . . But these imputed costs are not objective, but flat costs determined by the public policy of the State."20 This view, however, is easily refuted by recalling the very elements of price theory.

Why is there an objective price structure in a competitive market? Because, as a result of the parametric function of prices, there is generally only one set of prices which satisfies the objective equilibrium condition, i.e., equalizes demand and supply of each commodity. The same objective price structure can be obtained in a socialist economy if the parametric function of prices is adopted as an accounting rule. In a competitive market the parametric function of prices results from the number of competing individuals being too large to enable any one to influence prices by his own action. In a socialist economy, production and ownership of the productive resources outside of labor being centralized, the managers certainly can and do influence prices by their decisions. Therefore, the parametric function of prices must be imposed on them by the Central Planning Board as an accounting rule. All accounting has to be done as if prices were independent of the decisions taken. For purposes of accounting, prices must be treated as constant, as they are treated by entrepreneurs on a competitive market.

The technique of attaining this end is very simple: the Central Planning Board has to fix prices and see to it that all managers of plants, industries, and resources do their accounting on the basis of the prices fixed by the Central Planning Board, and not tolerate any use of other accounting. Once the parametric function of prices is adopted as an accounting rule, the price structure is established by the objective equilibrium condition. For each set of prices and consumers' incomes a definite amount of each commodity is supplied and demanded. Condition C determines the incomes of the consumers by the prices of the services of ultimate productive resources and the principles adopted for the distribution of the social dividend. With those principles given, prices alone are the variables determining the demand and supply of commodities.

The condition that the quantity demanded and supplied has to be equal for each commodity serves to select the equilibrium prices which alone assure the compatibility of all decisions taken. Any price different from the equilibrium price would show at the end of the accounting period a surplus or a shortage of the commodity in question. Thus the accounting prices in a socialist economy, far from being arbitrary, have quite the same objective character as the market prices in a regime of competition. Any mistake made by the Central Planning Board in fixing prices would announce itself in a very objective way—by a physical shortage or surplus of the quantity of the commodity or resources in question—and would have to be corrected in order to keep production running smoothly. Since there is generally only one set of prices which satisfies the objective equilibrium condition, both the prices of products and costs are uniquely determined.

Our study of the determination of equilibrium prices in a socialist economy has shown that the process of price determination is quite analogous to that in a competitive market. The Central Planning Board performs the functions of the market. It establishes the rules for combining factors of production and choosing the scale of output of a plant, for determining the output of an industry, for the allocation of resources, and for the parametric use of prices in accounting. Finally, it fixes the prices so as to balance the quantity supplied and demanded of each commodity. It follows that a substitution of planning for the functions of the market is quite possible and workable.


Problems of Choice and Decisionmaking

The last two selections (1.10 and 1.11) raised some issues concerning the market and planning. These issues will reappear throughout other parts of this book—especially in chapter 3 when we discuss market failure and optimal pricing. Before examining these issues in greater detail, we should simply reemphasize at this introductory stage that the economics of development does not exclude price. Neither demand nor supply is normally inelastic. On the contrary, the quantities demanded and supplied will change in response to a change in price. Knowledge of this responsiveness to price is indispensable for policymaking.

It will also be recalled that we began by interpreting economic problems as problems of choice. By recognizing that price is a coefficient of choice (Selection 1.9), we can realize that a price system is essential for dealing with the central problem of scarcity.

An analysis of the price system can be conducted, however, at two levels—as descriptive theory or as normative theory. Descriptive theory refers to actual behavior of markets and economies. Normative theory refers to optimal behavior—it prescribes what policy ought to be followed. The rational choice models developed by economists can be interpreted at both levels. The analysis developed to understand actual behavior is also helpful in determining desired behavior. A decisionmaker wants to make an optimal choice. An understanding of the concepts already introduced in this part should help the decisionmaker do this—an understanding of scarcity, competing ends, opportunity cost, efficiency, demand, supply, elasticity, surplus, production costs, economies and diseconomies of scale, and marginal analysis.

These concepts underlie the discussion in the final selection (1.12), which sets forth in summary fashion the model of choice from the viewpoint of a public decisionmaker.

1.12 The Model of Choice

An economist approaches a decision by asking “What do we want and what can we get?” Ordinarily we want more than we can get, and because our capabilities are limited and the resources available to us scarce, choices must be made among our competing desires. The Port Authority would like to expand airport operations and at the same time reduce noise levels. It cannot do both; as headlines testify, the choice is difficult. How choices should be made—the whole problem of allocating scarce resources among competing ends—is the stuff of economics and the subject of this book. We focus on choices in the public sector, on how decisions should be made by governments at all levels and by nonprofit institutions. As we are by now all well aware, the government is not a business, and in many respects it cannot be run like a business. Its goals are different and it operates under different constraints. Yet the basic elements of good decisions are the same in all arenas, and the methods for making them set forth here are applicable for all decisionmakers, public and private.

Our starting point is a fundamental model of choice, a model that those who have studied economics will doubtless find familiar. We have seen that a model is a simplified representation of some aspect of the real world, a deliberate distillation of reality to extract the essential features of a situation. The fundamental choice model is particularly valuable because it offers a universal yet succinct way of looking at problems in terms of the two primary elements of any act of choice:

1. The alternatives available to the decisionmaker; and
2. His preferences among these alternatives.

Moreover—and this is not to be taken lightly—the model forces the decisionmaker to express the alternatives he faces and his preferences among them in comparable units. You will see from our examples that the alternatives may sometimes be described in tangible terms, actual outputs that can be seen and counted, such as electricity and water, or allergy tests and electrocardiograms. At other times the outputs of the alternative choices will be described in terms of intangible attributes such as intelligence and beauty, or taste and nutrition, or safety and speed. Some of these intangibles can be measured more or less objectively; others cannot. The model is flexible; it easily handles all types of attributes, whether described by hard numbers or paragraphs of prose, so long as the decisionmaker's preferences are expressed in the same terms as the alternatives.

In this chapter we set forth the model of choice and deduce from it the characteristics of best decisions. We begin by looking at the model in drastically simplified situations, deliberately limited to two variables—two outputs or attributes—so that we can plot the results on a graph. Tradeoffs—painful tradeoffs—are the essence of difficult decisions, and there is no clearer way to visualize the nature of these tradeoffs than with a graph. In general, of course, many more than two variables will be considered by a policymaker; the concepts are readily extended to many variables, and the geometric representation to many dimensions.

Our perspective is that of the unitary decisionmaker, whether an individual or a corporate body, who faces no significant uncertainties. The model applies equally well to a person making a choice about his own consumption or career, or a firm making a production or research or marketing choice, or a public official facing a decision about a government project.

The Alternatives Available to the Decisionmaker

The first element of the basic model describes the alternatives available to the decisionmaker. If this were a standard economics text, we would introduce you to apples and oranges and ask you to consider the plight of the grocery shopper who must allocate his fruit budget between those two goods. But this is a book about public decisions, so we ask you instead to play the part of a public official who must choose among several alternative dam projects. These projects are identical in every respect—costs, environmental consequences, and so on—except two: they produce different amounts of electric power and water for irrigation. In other words, the decisionmaker faces a certain number of alternative quantities of power and water. Suppose there are five possibilities with the following outputs.

<table>
<thead>
<tr>
<th>Dam</th>
<th>Electric output (thousands of kilowatt-hours a day)</th>
<th>Water output (millions of gallons a day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>E</td>
<td>6</td>
<td>25</td>
</tr>
</tbody>
</table>

We plot these five alternatives on a graph, as shown in figure 1-27. In this diagram, point A represents an electric output of 22,000 kilowatt-hours a day and a water output of 20 million gallons of water a day.

Increases in both outputs would clearly be desired by the decisionmaker. (Remember that we have assumed that the projects are the same with respect to all other factors, including cost.) This means that any point (such as C) lying both north and east of another point (such as D) is superior to it. In technical language, C dominates D; that is, it is better in every respect. The decisionmaker would never choose D in preference to C regardless of his relative preferences for electricity and water. A combination of outputs or attributes is said to be efficient if, given the available alternatives, it is impossible to increase one output without giving up some of at least one other. In figure 1-27, points A, B, and C are efficient; D is not because we can have more of both water and electricity, at the same cost, by choosing C. Similarly, E is not efficient because we can have more of both water and electricity by choosing either B or C. Obviously, dominated points can never be efficient; thus all points are either dominated or efficient.21

21. You may come across the expression "X weakly dominates Y." This means that X is equally as good as Y in some respects, better in at least one, and worse in none.
It need not be the case that the decisionmaker has only a small number of discrete alternatives open to him. He might well be able to vary the output combinations in small steps so that he could always get a little more hydroelectric power by giving up a little more water for irrigation. If it is possible to make these continuous tradeoffs we are not limited to five or eight or even twenty-five separate points; we can draw a whole curve that delineates the location of efficient points. This curve (shown in figure 1-28 as EW) tells us the maximum achievable output of water for every possible output of electricity. For example, point F indicates that with an electric output of 8 thousand kilowatt-hours, the maximum water output is 32 million gallons, and conversely. We call this collection of efficient points the possibility frontier.

The possibility frontier may be applied in numerous contexts. It may describe tangible goods or intangibles. It may refer to a public project, as here, or to a private firm or individual. It may be used to describe the combinations of goods available to a consumer or the tradeoff of intangible attributes for a proposed public building, for example flexibility of space and privacy of offices. The frontier may be straight or curved, continuous or discrete, or it may consist merely of a few isolated points. Whatever shape it takes, it is the set of efficient alternatives, the contenders for the best available choice.

The following examples illustrate the wide applicability of the possibility frontier:

1. A transit authority has a $10 million budget for new rolling stock. Buses cost $65,000 each and subway cars $465,000. The efficient choices for that budget are indicated by its possibility frontier, line CB in figure 1-29. The authority would never spend $10 million for interior combinations even though they are feasible.

A similar straight-line possibility frontier is appropriate for a municipal highway department that can readily transfer its men and equipment from street repairs to snowplowing as the need arises.

In a more familiar vein, a straight line sets forth the maximum combinations of apples and oranges that the consumer can buy, at going prices, for a given sum of money. A frontier of this type is frequently termed a budget line.

2. An urban renewal project includes a high-rise building. The lower floors will be rented as commercial space; the upper floors will be used for apartments. It is undecided where the line should be drawn, although certainly no more than five floors will be rented commercially. The tradeoff between the value of the commercial rentals and the amount of housing space available is shown by the curve CR in figure 1-30. The curvature reflects the fact that because of the escalator and elevator configuration, commercial space is less valuable as you go higher in the building.

If the municipal highway department's maintenance and plowing equipment were less than fully flexible, a tradeoff curve similar in shape to this one would result. If all equipment is being used for maintenance when a snowstorm strikes, the trucks diverted first to plowing should be those that will pro-

22. Some readers will recognize that this particular curve is in fact the "production possibility frontier" talked about so frequently by economists. We deliberately drop the word "production" because we use the term "possibility frontier" in a somewhat broader sense.

23. This possibility frontier is simply the straight line represented by the equation $65,000B + 465,000S = 10,000,000$, where $B$ is the number of buses purchased and $S$ is the number of subway cars. You may wonder about the smooth line and the problem of fractional buses and subway cars. Under most circumstances, when outputs are indivisible, the best choice indicated by the model will be only approximate; the basic principles still hold.
duce the most miles plowed relative to the maintenance capability sacrificed. If the storm is a really bad one, trucks that are less satisfactory will have to be pressed into service, at a greater maintenance sacrifice.

3. A legal aid office provides two kinds of services to its clients, “domestic relations” and “landlord-tenant.” The maximum combinations of cases are indicated by the curve DL in figure 1-31. This sort of curvature would arise if the average time required to handle a type of case shrinks as the legal staff handles more and more cases of that type.

4. Three sites (1, 2, and 3) are available for a new health clinic. Each is rated A, B, C, or D with respect to the two characteristics deemed relevant, accessibility to the community and quiet. The possibility frontier, which is shown in figure 1-32, consists of three points. (This figure, incidentally, shows us a common way of constructing a diagram when only letter grades or ordinal rankings—such as first, second, and so on—are available. The letters or numbers run “backward” from the origin because we are accustomed to seeing things get better as we move north and east.)

In short, the set of feasible alternatives may be specified in many different ways. The efficient choices are those from which it is not possible to make an improvement in one respect without accepting a sacrifice in some other respect.

The Decisionmaker’s Preferences

The second element of the fundamental choice model describes the decisionmaker’s preferences. We assume throughout the technical part of this book that these preferences are given, though occasionally we point out where a particular technique is helpful in defining preferences.

We have seen that frequently the available alternatives combine many outputs or attributes, all of which are valued by the decisionmaker. The problem of choice is not difficult if one alternative is superior to all others with respect to all attributes—in other words, if one alternative is dominant. Of the five dam projects considered above, C dominates D, and B dominates E, but no single alternative is dominant over all the others. We can eliminate D and E, but there is still more than one efficient alternative.

Unfortunately, we are rarely presented with a dominant choice. Consider the case of three applicants for a job. Anderson, Barker, and Corcoran, who are equally qualified in all attributes but two, technical capability and ability to get things done. Their rankings with respect to these two attributes are summarized in the following table. Here there is no dominant choice, for the applicant who is the best technician of the three is not best in ability to get things done. An employer might even prefer Barker, who is neither the best technician nor the most able, to either Anderson or Corcoran.

<table>
<thead>
<tr>
<th>Attribute rankings</th>
<th>Applicants for a position</th>
<th>Technical capability</th>
<th>Ability to get things done</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson</td>
<td>1st</td>
<td>3rd</td>
<td></td>
</tr>
<tr>
<td>Barker</td>
<td>2nd</td>
<td>2nd</td>
<td></td>
</tr>
<tr>
<td>Corcoran</td>
<td>3rd</td>
<td>1st</td>
<td></td>
</tr>
</tbody>
</table>

Such conflicts in attribute rankings lie at the core of many difficult decisions that must be made in the
public sector. These conflicts may arise because the decisions affect many people; although policy A is better for one group in the society, policy B is better for another; a battle over an airport location is an example. How would you weight the attributes "better for group 1" and "better for group 2"? Or if time is a crucial element we may find, for example, that policy A (screening for bladder cancer) is more immediately beneficial, but that policy B (cancer research) will be better for a long period starting 20 years from now. How would you weight the attributes "better now" and "better later"? In a third context, policy A will be superior if some uncertain events turn out favorably, but policy B—flood plain zoning, for instance—is a better hedge against misfortune. Again, how would you weight the attributes "better if no flood" and "better if flood"? Conceptually, these conflicts among attributes are no different from the conflict between the attribute rankings of Anderson, Barker, and Corcoran. What is needed in all these cases is a method for determining and displaying the decisionmaker's preferences among different combinations of attributes.

If we develop a perfectly general formulation that applies to all combinations, we can then apply it to some of the more complex problems we will encounter later on.

As in the last section, we will restrict our discussion to situations in which there are only two relevant attributes, because this permits graphing. The ideas may be extended readily to many dimensions.

Let's get back to our dam planner, whom we left contemplating the five projects and wondering what to do. Forget, for the moment, about the particular dams in question and think only about his unrestricted preferences for electricity and water. Suppose we consider just one combination of outputs, say 11 thousand kilowatt-hours of electricity per day and 25 million gallons of water per day. This combination is shown at point P in figure 1-33. The model presumes that as far as the decisionmaker is concerned there are other combinations of electricity and water that are just as good—and no better. Say he is neutral between points P and Q, where Q represents 7 thousand kilowatt-hours per day and 36 million gallons of water per day. And he regards P and Q as exactly as satisfactory as R (19 thousand kilowatt-hours and 17 million gallons), and S (24 and 14). We draw a curve, 1, in figure 1-33, through these four points and all others that he regards as equally desirable; we call it an indifference curve. Note that we say nothing about the amount of satisfaction depicted by this curve. We state only that all points on it are equally good from the decisionmaker's point of view; that is why we have drawn it as we have. This causal relationship is crucial; it is not because they lie on the same curve that the points on an indifference curve are equally satisfactory—rather, they lie on the same indifference curve because they are equally satisfactory. Note further that when we draw a smooth curve such as 1, we in effect assume that the decisionmaker's preferences are continuous.

Similarly, we could draw other indifference curves 0, 2, and 3, depicting lower and higher levels of satisfaction, respectively, as shown in figure 1-34.

Again, we do not assign specific values to these levels of satisfaction; we merely state that 1 is better than 0, 2 than 1, and 3 than 2. In particular, there is no implication that movements of equal distance across the graph are equally valuable. We could draw more and more indifference curves, until there is one (and only one) through every point in electricity-water "space." As in the diagrams of available alternatives, things get better as we move north and east. The decisionmaker's preferences for possible combinations of electricity and water are thus com-

Figure 1-33. Indifference Curve on Which All Points Are Equally Satisfactory
not build fractional courts or diamonds, just as a transit authority would not buy fractional buses or subway cars. Each contour really consists of a succession of discrete points.)

3. An individual is a swimming and theater enthusiast; he likes to spend some time doing each. The more he swims, the more theater time he will give up for still another hour of swimming. His preferences are shown in figure 1–37. $I_2$ is preferred to $I_1$, and so on.

4. A hospital outpatient clinic provides two types of medical services, allergy tests and electrocardiograms. The directors of the clinic originally considered three possible allocations of space and staff, which would permit the combinations of services shown in the following table.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>EKGs a day</th>
<th>Allergy tests a day</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>65</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>40</td>
<td>65</td>
</tr>
</tbody>
</table>

not build fractional courts or diamonds, just as a transit authority would not buy fractional buses or subway cars. Each contour really consists of a succession of discrete points.)

3. An individual is a swimming and theater enthusiast; he likes to spend some time doing each. The more he swims, the more theater time he will give up for still another hour of swimming. His preferences are shown in figure 1–37. $I_2$ is preferred to $I_1$, and so on.

4. A hospital outpatient clinic provides two types of medical services, allergy tests and electrocardiograms. The directors of the clinic originally considered three possible allocations of space and staff, which would permit the combinations of services shown in the following table.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>EKGs a day</th>
<th>Allergy tests a day</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>65</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>40</td>
<td>65</td>
</tr>
</tbody>
</table>

Figure 1-36. Preference Function Based on Maximum Simultaneous Use of Facilities

Figure 1-35. Preference Function Based on Sales of Thermometers and Barometers

Figure 1-34. Indifference Curves Showing Different Levels of Satisfaction
The directors concluded that of these B was the most desirable and A the least desirable. They are now given the opportunity to consider additional mixes of services. What can we say about their preference function, on the basis of the rankings already provided? Strictly speaking, we know only that point B is preferred to C, which in turn is preferred to A. This suggests that the indifference map bears some resemblance to that shown in figure 1-38, but we can be sure only of the ordering indicated.

In sum, there is a wide range of situations in which we can model the decisionmaker's preferences. In the next section we consider how preferences and alternatives can be put together to define the best choice.

Figure 1-38. Preference Function of Alternative Mixes of Services

The Best Choice

We have now represented on a geometric diagram the two elements in the act of choice, the set of alternatives from which the choice is to be made and the set of preferences according to which the chooser ranks these alternatives. What use can we make of these representations of options and preferences? Because both are measured in the same units, which in the dam example were kilowatt-hours of electricity per day and gallons of water per day, we can plot both on the same graph. (This is a very important point; we'll get back to it in a minute.) The nature of the best choice is then indicated. If the alternatives are limited to the five specific dams originally considered, our diagram will look like figure 1-39, which combines figures 1-27 and 1-34. The policy planner should choose alternative C, for it is the project preferred according to his own ranking; it is the project that places him on the highest indifference curve.

By contrast, figure 1-40 shows that a different indifference map might well produce a different ranking of the alternatives. (We use \( J_s \) to label this set of indifference curves.) Now B is preferred to C. Of course no map would ever show D or E as the best choice. Figure 1-40, incidentally, shows a decision-maker who really likes his water; it takes a lot of electricity to persuade him to give up a little water.

Figures 1-39 and 1-40 demonstrate the preferred choice when the decisionmaker confronts discrete alternatives. What happens when continuous trade-offs between electricity and water are possible, as in figure 1-28? Figure 1-41 combines the continuous...
Figure 1-40. Indifference Map Based on Preference for Water and Electricity

Figure 1-41. Point of Optimum Choice Using Possibility Frontier and Indifference Map

possibility frontier of figure 1-28 and the indifference map of figure 1-34. Examining this composite diagram, we see that the best choice for the dam planner is the combination of electricity and water represented by point T, because only at that point can he reach the highest possible indifference curve. A rational decisionmaker would never choose a point inside the frontier—an inefficient point—because larger outputs of both attributes would be attainable at points on the frontier at no additional cost. And he would not choose a different point on the frontier (F, for example) because, as his indifference map tells us, all the points along the FG segment of the frontier are better for him than F. He would like to choose a point on I₁, but he can't—he can't move beyond the possibility frontier. Note very carefully the relationship between the possibility frontier and the indifference map: they are bulging toward each other, and at point T they just touch. That is, at that point the possibility frontier is tangent to an indifference curve; their slopes are equal.

Before reading on, you should convince yourself that the point of tangency is indeed the best choice in this diagram, and that it will always be the best choice if the curves have the general shape shown in the diagram.²⁴

We noted that at point T the possibility frontier is tangent to the planner's indifference curve, and the slopes of these two curves are equal. This is a geometric representation of an important characteristic of the optimum choice. Before discussing the significance of this equality for the model of choice, we must interpret the meaning of these slopes.

Look again at the possibility frontier EW; it is reproduced in figure 1-42. As we move down along this curve from E toward W, we are in effect reducing the output of electricity in return for an increased production of water. The rate at which this tradeoff is taking place is called, quite reasonably, the rate of transformation, and the possibility frontier is sometimes called the transformation curve. For the possibility frontier we have assumed, points G and F have the numerical values shown in the following table.

<table>
<thead>
<tr>
<th>Point</th>
<th>Electricity (thousands of kilowatt-hours a day)</th>
<th>Water (millions of gallons a day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>Change</td>
<td>7</td>
<td>12</td>
</tr>
</tbody>
</table>

Over the range from G to F, we say that the rate of transformation of electricity for water is 7,000 kilowatt-hours for 12 million gallons or 1 kilowatt-hour for 1,714 gallons.²⁵ We are dealing, however, with transformation possibilities that change continuously as we move along the curve. Although we

²⁴. It may occur to you that we require certain assumptions about the shape of the curves if this statement is to hold. You are quite right. Fortunately the assumptions are ordinarily plausible; they are discussed in the next section.

²⁵. To be strictly accurate, the rate of transformation is a negative number, for the changes in the two variables are of opposite sign as we move along the curve. Since the essential idea of this model is tradeoffs, tradeoffs that involve less of one thing in return for more of something else, all these rates would have minus signs if we insisted on rigor. In other words, all the curves slope down to the right. Because the slopes are all negative, it is customary to drop the minus sign in verbal discussions such as this.
can average only 1,714 units of increased water production for every kilowatt-hour of electricity sacrificed over the entire GF range, we would have to sacrifice less than 1 kilowatt-hour of electricity to get the first 1,714 additional gallons of water at point G. (Analogously, while a man may drive from Chicago to Denver at an average speed of 50 miles an hour, part of the time he is going faster than 50—say, 60—and part of the time he is going more slowly.) At any point on the curve, the rate at which one output can be transformed into another is given by the slope at that point on the possibility frontier. The steeper slope at point F indicates that the rate of transformation at which electricity can be traded for water is greater there than it is at point G. We refer to the rate at which one output can be transformed into the other at a particular point as the marginal rate of transformation, or MRT. The use of the word marginal suggests that we are considering only a very small region, and that we should be alert to the possibility that the MRT will change if we move a significant distance along the curve. For the individual consumer who has no influence over the prices of the things he buys, regardless of how much he buys the MRT is constant. Hence his budget line is a straight line.

The slope of an indifference curve may be interpreted in similar fashion. It represents the way in which the decisionmaker is willing to trade electricity for water while still remaining at the same level of satisfaction, in other words, on the same indifference curve. The steepness of the indifference curve indicates the rate at which he is willing to trade off between the two outputs; the steeper the curve, the greater the amount of electricity output he is willing to trade for a unit of water production. Thus in figure 1-43 he would give up more electricity at point M in return for an extra unit of water than he would at point N. We refer to this tradeoff rate as the individual's rate of substitution between the two goods: the tradeoff rate at a particular point is called the marginal rate of substitution or MRS. It is a logical complement to the marginal rate of transformation or MRT.

The relation between the marginal rate of transformation and the marginal rate of substitution is an important one to grasp. The MRT is the rate at which one is able to exchange one good for another at a particular point; the MRS is the rate at which the decisionmaker is willing to exchange one good for another. The former is defined by the typical constraints of the production process, the latter by the subjective preferences of the decisionmaker. The preferences of the public decisionmaker, to be sure, should not be his own, for he is not the consumer. Rather they should reflect his view of the preferences of the society he represents. The way social preferences should be determined is no easy matter.

We saw that for the optimal choice the slopes of the possibility frontier and the indifference curve are equal, because the two curves are tangent at that point. Thus at the best point the marginal rate of transformation between two goods must equal the decisionmaker's marginal rate of substitution between them. The objective rate at which the planner is capable of carrying out the tradeoff between electricity and water power is exactly equal to the subjective rate at which he is willing to substitute between these two outputs. Why must this be the case? Consider point G in figure 1-44, at which the rate of transformation between electricity and water (the rate at which electricity output can be traded for water output) is less than the decisionmaker's rate of substitution. He will find it to his advantage to pursue
Figure 1-44. Optimal Choice Occurs at Point Where Marginal Rate of Transformation and Marginal Rate of Substitution Coincide

![Diagram showing optimal choice at point where marginal rates coincide]

options that yield more water and less electricity. At G, he is able to get 2,200 gallons of water output by giving up 1 kilowatt-hour of electricity, whereas he would be willing to sacrifice 1 kilowatt-hour for as little as 1,050 additional gallons of water. Clearly, then, it is in his interest to trade off electricity for water at the advantageous rate available at point G. In this manner, the decisionmaker will continue to move along the possibility frontier until he reaches point T, where the willingness rate and the possibility rate are equal. If he strays beyond T, this same type of marginal analysis will lead him back to the point of tangency.

We have seen in our example that the model of choice may be interpreted in various ways. The chooser may be an individual, a group operating as a unit, or a government. The alternatives may be simple combinations of goods or outputs like those we've considered here, or they may be far more complex. Uncertain prospects such as alternative lottery tickets, or the choice among elaborate plans for an individual's entire life, or decisions about sweeping programs of social reform are all encompassed by the model. The principle of rational choice is unchanged. The decisionmaker should attempt to select the available alternative that he values most highly in terms of his own preferences.

MARGINAL ANALYSIS

This discussion of marginal rates of transformation and substitution is only one example of the type of analysis that forms the core of traditional microeconomic theory. In a nutshell, in order to achieve an optimal result, the allocation of scarce resources among competing uses must satisfy certain marginal equalities. For example, the consumer should allocate his budget so that he gets the same satisfaction from the last dollar he spends on orange juice and the last dollar he spends on going to the ballet. And a rational consumer will do just that, even though he will rarely do so consciously. A farmer or the manager of a pencil factory should expand production just to the point where his last dollar of sales costs him exactly $1. Producing more diminishes his profit, producing less means that he forgoes some of the profit he might have reaped. Similarly, a public decisionmaker—a mayor, say—should allocate spending on park maintenance and on fire protection so that the last dollar spent on each is equally satisfying to the society he represents.

Marginal analysis is conventionally expounded in terms of the private sector. Because scores of texts in microeconomic theory cover the subject admirably, we will not delve into its details here.

MORE ON DIMENSIONS

The model of choice requires that preferences be expressed in the same units as the outcomes of the various alternatives proposed. Thus, if the decisionmaker is offered a choice among assorted combinations of apples and oranges, his preferences must be expressed also in terms of apples and oranges. Conversely, if he is to choose a mix of strange fruit whose attributes are a mystery to him, although he knows his preferences for, say, vitamins and juiciness, the outcomes of the various possible choices must be expressed not as bundles of fruit but as combinations of these attributes. In other words, he must be able to measure these fruits in terms of the characteristics he understands, cares about, and can work with.

For most of us, apples and oranges are familiar goods; if we select a given bundle, we know what we are getting and we can think directly in terms of our preferences for its contents. In contrast, the outcomes of policy choices are rarely obvious and frequently subtle. In considering a proposal made for a new national park on Boston's harbor islands, the knowledge that the choice is between developing the islands and leaving them as they are tells us little about the ultimate outcomes of these two courses, for the bare choice between developing and not developing is almost meaningless. It is the impact of that choice on recreation, the environment, the local economy, and so on, that we care about. In the case of the dam proposals, the planner's preferences for electricity and water presumably depend on the uses to which these outputs will be put. That is to say, his true preferences relate to such considerations as the number of homes served by the electricity and the number of acres irrigated by the water. But unless he can translate these preferences into kilowatt-hours of electricity and gallons of water, he can't determine which dam is the best choice. Alternatively, of course,
he could have expressed the output of each dam in terms of homes served and acres irrigated.

This matter of dimensions may seem embarrassingly obvious, but it is not trivial. We are all familiar with controversies about ongoing public programs—public housing, for instance—in which the proponents point with pride to admirable intangible goals while the opponents add up the broken windows and the muggings in the elevators. In short, being forced into a model of choice is useful, for it compels us to get the ground rules straight on what we value and how the alternatives before us measure up in these respects.

**What Practical Use for the Model of Choice?**

Of what use is the model of choice to the decisionmaker? Is not the ability to make difficult choices among competing ends the very asset that has placed him in his policymaking position? The economist would answer that systematic analyses and formal statements of procedure can be valuable. He might invoke an analogy to sailing and the theories that underlie it: even the experienced yachtsman benefits from a knowledge of aerodynamics.

For expository purposes, we ordinarily describe the fundamental model in terms of simple choices, so simple that they can easily be made without recourse to analytic methods. (After all, we need neither a theory nor a model to choose among ice cream cones of different flavors, although we would argue that implicit models do in fact govern such choices.) Moreover, the model of choice assumes an individual decisionmaker who is confronted with a fixed menu of alternatives in a world he can't change. He faces no significant uncertainties, and he has complete command over his preferences. These are stringent conditions: they are unlikely to be fulfilled precisely in most real-world situations. Nevertheless, it is important to understand the framework that produces the best decision in a well-behaved situation. This model can serve as the foundation on which to build choice procedures for more complex situations.

What new complications arise in public policy choices? At the outset, the mere process of constructing the possibility frontier and the preference function is vastly more difficult. The choice between allocating $100 million to job training programs and using it for bolstering Title 1 grants for the education of disadvantaged children may not only plumb to the roots of our value system, it will involve us in difficult problems of predicting the costs and benefits that each will generate for quite different groups of people. Even in the evaluation of individual projects we may encounter exceptional difficulties. How can we weigh the losses of those families displaced by our hypothetical urban renewal project against the gains of those who will enjoy improved living conditions in the completed housing? How can we relate the basic model to decisionmaking in the real world, to issues such as how to allocate funds for biomedical research or how to regulate automobile safety equipment?

Our approach is to extend the fundamental model by employing it as a prescriptive model for analyzing and making such complex decisions. Note carefully that this is the converse of the descriptive use of the model in economic theory. The economist would construct the possibility frontier and the decisionmaker's preference map by observing the choices that can be made and that actually are made, and would then argue that the decisionmaker behaves as if he were following such a model. The economic model is thus interpreted as a descriptive theory of behavior, a theory that can be tested by seeing whether actual choices correspond to those predicted by the model. We ask our decisionmaker to invert this process. We expect him to think about and determine, consciously and explicitly, what the alternatives and their outcomes are and what his preferences among these outcomes are, and to make his choice accordingly.
CHAPTER 2
Functions of Prices

Having considered the structural elements of a price system, we may now examine the importance of prices more directly by analyzing the most important functions of prices: information, allocation, rationing, mobilization, and distribution.

Informational Function

By providing market signals, a price system is an inexpensive mechanism to guide economic agents in decisionmaking. Because a price is a substitution ratio, indicating the rate at which one thing can be exchanged for another, it facilitates a decision. The decisionmaker need not incur the cost of searching for alternatives. Moreover, when conditions of demand or supply change, the changes are reflected in price changes that alter the responsiveness of buyers and sellers. The buyer or seller can act on the price change without probing into its cause. For example, a price rise might be the result of a harvest failure, but the buyer need not know the cause to reduce consumption of the affected product. It is sufficient just to be aware of the price increase. The informational function of prices is crucial to the operation of a decentralized economy in which economic decisions are made by individual producers and consumers. Although each individual operates on limited information about the entire economy, the multitude of individual production and consumption decisions is harmonized by the price system. The significance of this informational efficiency is emphasized in the following selections.


2.1 Limited Knowledge and Economic Analysis

Let us return to the optimization problem of the individual. One aspect on which we put a good deal of weight, particularly in our less formal discussions, is that a market system is informationally economical. That is, we tend to regard it as a virtue of the system that the individual agent need not know very much. Specifically, he is supposed to know the motivation and production conditions which define him, i.e., his utility function and production possibility set, together with the prices of the commodities he buys and sells. The economic system, taken as a whole, has vastly more in it than any one individual knows; it contains the utility functions and production possibilities of all individual agents. Indeed, the apparent modesty of the information needed is one of the most appealing aspects of the neoclassical model, both in the descriptive sense that the individual’s decision problems appear manageable for him and for the economist studying him, and in the normative sense that the system permits its members to spend their time and effort at producing goods rather than in unnecessary duplication of information.

But clearly this simplification of the individual’s decisionmaking is made possible only because the markets have supplied the information economized on, in the form of prices. In equilibrium, at least, the system as a whole gives the impression of great economy in the handling of information, presumably because transmission of prices is in some significant sense much cheaper than transmission of the whole set of production possibilities and utility functions.
It is this point which emerged in the great debate over the feasibility of socialism begun by Ludwig von Mises's attack and usually thought of as concluding with the work of Oskar Lange and Abba Lerner in the 1930s; though it should be added many of the essential points had already been made earlier by Vilfredo Pareto and Enrico Barone. What was argued, in effect, was that a socialist system could use the price system and therefore achieve whatever economies in information it does achieve; and if the equilibrium conditions are written out they do give the appearance of relative simplicity. But what was left obscure is a more definite measure of information and its costs, in terms of which it would be possible to assert the superiority of the price system over a centralized alternative. Though I feel that current work has brought about a considerable clarification, we still have no definite measure. Indeed, in some respects, more recent developments have made the answers less clear. Several writers, in both Western and socialist countries, have noted that alternative decentralized schemes exist where quantity messages rather than price messages are transmitted in the successive stages of approximation and that such schemes also have efficient equilibrium points. Indeed, with the development of mathematical programming and high-speed computers, the centralized alternative no longer appears preposterous. After all, it would appear that one could mimic the workings of a decentralized system by an appropriately chosen centralized algorithm. While there is more to the story than these few remarks, they do make the point that if we are going to take informational economy seriously, we have to add to our usual economic calculations an appropriate measure of the costs of information gathering and transmission.

But actually the comparisons between socialist and capitalist resource allocation systems have tended to overlook some of the most obvious facts while examining finer points closely. As we all know, both production and consumption decisions are in fact made with reference to the future as well as to the present. A rational production plan includes very importantly decisions or at least plans about the future; and similarly with consumption plans. Investment and savings not only are integral parts of our current decisions but in the long run also shape the possibilities for further development. As we know, the formal neoclassical model can be extended to decisions over time by dating commodities and regarding the same commodity at different dates as different commodities. All previous conclusions follow; allocative efficiency, for example, is achieved with the same appearance of informational efficiency.

But of course there is a slight problem with this reasoning. The information about future commodities needed includes their prices. These prices must be those found on a suitable market, one in which future supply and future demand are equated. Unfortunately, no such markets exist. Even the futures markets in certain commodities, limited in extent as they are, do not in fact lead to balancing all future decisions. Rather they balance present commitments to the future; but it is understood by all parties that when the future becomes the present, there will be a spot market on which the futures commitments may be undone; and indeed those making no futures commitments at all can enter and know now that they will be able to enter.

Even as a graduate student, I was somewhat surprised at the emphasis on static allocative efficiency by market socialists, when the nonexistence of markets for future goods under capitalism seemed to me a much more obvious target.

However that may be, the nonexistence of these markets must be faced. Now in general equilibrium any part of the system affects every other part in at least two different ways. Thus, we may ask two questions about the nonexistence of futures goods markets: what are its implications for the rest of the system and what are the reasons for its nonexistence? The implication first of all is that the information needed by the optimizer is not provided by an existing market. It will be provided by a market which will exist in the future, but that is a bit too late to help in decisions made today. Hence, the optimizer must replace the market commitment to buy or sell at given terms by expectations: expectations of prices and expectations of quantities to be bought or sold. But he cannot know the future. Hence, unless he deludes himself, he must know that both sets of expectations may be wrong. In short, the absence of the market implies that the optimizer faces a world of uncertainty.

2.2 Market and Planning

Two Extreme Views

There are two extreme views of the role of the market in the functioning of economic systems. One proposes a “pure” market economy; the other, “pure” central planning.

According to the former the market is in itself capable of controlling the economic system. Insofar as the price system meets the theoretically established requirements of optimality, it in itself will provide

the basic information necessary for control. Any interference with the economic processes is superfluous.

According to the latter view, the economy must be centrally regulated according to plan. If planning is of a sufficiently high standard, exact and reliable, there is no need for any other control; in particular, it would be superfluous to expose the economy to fluctuations and frictions from the market.

A justification of either view could be proven only with extremely strong abstractions foreign to reality. It is paradoxical but true that we would have to make essentially the same unrealistic assumptions to "prove" the justification of either of the extreme views. Among other things, we would accept the following:

1. There is strict rationality; the "homo economicus" predominates in the economic system, whether on the lower level (according to general equilibrium theory), or on the higher (according to the theory of perfect planning).

2. There is no uncertainty in the economic system. The consequences of every decision can be foreseen.

3. In addition, the mathematical models which formulate the "perfect market" or the "perfect central planning" are compelled to apply strong restricting assumptions about the real sphere (for example, elimination of increasing returns and convexity of the set of production alternatives).

The problem is that none of the above assumptions is acceptable. (Each has been treated elsewhere several times and thus there is no need for refutation in this place.) To ask, "Planning or market"—is to ask the wrong question. Rather, what we must deal with are two, complementary control subsystems of the complicated and complex economic system.

Comparison of the Two Subsystems

We previously introduced the notion of control subsystem to denote the relatively separate parts within the control sphere, and we listed five kinds of them in all. From among the five we deal here with two, the market and national economic planning.

To make the survey easier we present the comparison in tabular form. Table 2-1 compares the information flows in the two subsystems. Both subsystems fulfill a useful role, but neither is fully reliable.

From the table, it can be seen that we interpret the concept of market broadly indeed. There are markets where the buyer and the seller freely agree on the price of the product. But we also use the word "market" for control subsystems regulating the market where the prices emerge in another manner, for example, where they are set by official authority or dictated by monopolistic firms. The market differs from other control subsystems in that the seller and the buyer are in immediate informative contact with each other and agree to the transactions with a "horizontal" information flow. In this sense there is no modern economy without a market and thus there was a market in socialist industry even at the times of highest centralization. The question is not whether there is a market or there is not (there must be one), but what kind of a market? What is the algorithm of price formation? Or, a question of equal importance: does pressure or suction prevail in the market? And what is the relation between the market and other control subsystems; what is its relative weight in the whole system of control?

Using table 2-1 let us now compare the two subsystems: the market and national economic planning.

The market works with fresh, actual information but it is short-sighted. Planning looks far ahead, but, accordingly, its data base is highly uncertain even with the most careful data collection.

The advantage of the market is that since the buyer pays money and the seller receives money, both of them thoroughly consider their offers and whether they should enter into contract. In this sense the information supply is responsible. The responsibility of the persons taking part in planning is several times removed. Their actions do not directly affect "their pockets." Therefore, irresponsible information is not infrequent. On the other hand, they are less biased, more objective. Buyers and sellers are necessarily egoistic; they consider primarily their own momentary interests. The two control subsystems differ from each other in their adaptive qualities.

The market is an adaptive, learning system where participants continuously grow wiser from their own earlier failures. For example, if a productive firm produces too much, it has trouble selling it all; if it has turned out too little, it forgoes good selling possibilities. There is a high cost of such failures; in the final analysis the market is blamed not only by the firm but also by society. Planning is cheaper. Plans are formulated by applying "trial and error" methods; the repeated trials take place only on paper or in "plan-bargains," not in the fluctuations of real processes. The adaptation cost that can be saved by planning is a gain for society. With careful planning it is possible to prevent disproportions which could be eliminated by the market only ulteriorly, at the cost of adaptation sacrifices resulting from fluctuations in real processes.

In the final analysis, we can draw the following general conclusions:
Neither the market nor planning can reliably control the modern, complex economic system alone. In itself, either one is a regulator which functions in less than a completely reliable way. Therefore, on the basis of the principle of multiplying information, the combined activity of the two is necessary for a satisfactory control of the system, for the improvement of its performance.

Factors Determining the Combination of Market and Planning

The above statement is of descriptive, real-science character; it establishes the existence of a historical tendency toward mixed economies. Every modern economic system is “mixed” in the sense that both subsystems appear in it.

But the statement says nothing concrete about the actual combination realized in some system, which depends on many factors. In the following we try to group the factors according to their major criteria.

1. The political, power, and ownership relations of the system. Under socialist order, social ownership relations play the dominant role and this promotes planning. On the other hand, under capitalist order private ownership plays the dominant role and this hinders planning. It is a historical fact that economy-wide planning began in the Soviet Union.

Private ownership is accompanied by competition, trade secrets, and the like, and these render general exchange of information difficult. A private firm does not like central interference in its own affairs. Of course, this attitude immensely hinders central planning. This attitude generates the illusion, spread partly under the effect of theoretical economists, that there is no need for central planning since the market can solve every control problem.

Analogous illusions are found within the socialist countries: illusions about the omnipotence of planning, about the unfailing foresight of economic processes and the undisturbed possibility of their rational control.

Thus, it was not only the actual power and ownership relations but also illusions and misbeliefs that developed under each of the two systems that acted

Table 2-1. Information Flow in the Market and the Planning Subsystems

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Market</th>
<th>Planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major information types</td>
<td>Offers and counteroffers</td>
<td>Statistical reports</td>
</tr>
<tr>
<td></td>
<td>Advertising</td>
<td>Plan proposal and counterproposal</td>
</tr>
<tr>
<td></td>
<td>Order</td>
<td>Plan-bargaining, critique of the plan-proposal</td>
</tr>
<tr>
<td></td>
<td>Confirmation</td>
<td>Plan-decision</td>
</tr>
<tr>
<td></td>
<td>Contract</td>
<td>Instruction or recommendation to those implementing the plan</td>
</tr>
<tr>
<td></td>
<td>Modification of contract</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Payment for fulfillment of contract</td>
<td></td>
</tr>
<tr>
<td>Character of reflection</td>
<td>The information directly reflects the real action</td>
<td>The information reflects the real action perhaps through several transmissions</td>
</tr>
<tr>
<td>Time lag between anterior information and real action</td>
<td>Preceding it only slightly, almost simultaneous</td>
<td>Preceding it by considerable time (1–5–15–20 years)</td>
</tr>
<tr>
<td>Role of memory</td>
<td>Small; short-time horizon</td>
<td>Large; long-time horizon</td>
</tr>
<tr>
<td>Measure</td>
<td>Variables measured in both physical and value terms; price information has the outstanding role</td>
<td>Variables measured in both physical and value terms; the former have the outstanding role</td>
</tr>
<tr>
<td>Is the flow of goods accompanied by money flow?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Vertical versus horizontal flow of information</td>
<td>Both, but the horizontal flows are dominant</td>
<td>Both, but the vertical flows are dominant</td>
</tr>
</tbody>
</table>
to produce a one-sided emphasis on planning in the socialist countries and on the market in the capitalist ones.

One can recognize an entire economic system as capitalist or socialist according to the political, power, and ownership relations dominating in it. But within an economy, the ownership relations are usually not uniform. In the capitalist countries there exist state-owned firms as well as firms owned by social institutions; and there are means of production under private ownership in the socialist countries. Accordingly, the concrete proportions between public and private property also influence the combination between planning and market. A broadening of the scope of public property will act in the direction of increasing the role of the planning subsystem.

2. General economic policy and, within that, the policy of raising the volume of production and the standard of living. The more general economic policy is directed toward increasing the volume of real processes, raising the growth rate, the more does planning come to the fore. Planning promotes a high concentration of resources in those actions which immediately promote the growth of volume. This happened in the Soviet Union and in other socialist countries. But also this is the situation in the developing African and Asian countries, which set themselves the task of rapidly liquidating backwardness. A highly centralized planning has appeared in these countries as well.

When the consumption by the population is on a low level and stagnates on that level (or rises only slowly in comparison to that level), it is relatively easy to plan centrally the production of consumer goods. The more the planning of consumers’ needs comes into the foreground the more it becomes necessary for production to adapt to consumers’ needs and for the production of consumer goods to become increasingly differentiated. Increasingly it is difficult to plan everything centrally. This explains the division of emphasis over the last fifteen years in the socialist countries between raising the level of living more quickly and subjecting the methods of economic administration to reform.

3. Pressure or suction. Pressure strengthens the role of the market, and suction that of planning (although either general state of the market may exist under either subsystem). In the case of suction, “rationing” is necessarily introduced. On the other hand, in the case of pressure, the control of most real processes can be left to a direct agreement between interested sellers and buyers. This is one of its major advantages.

Factors 1-3 hitherto listed affect the whole of the economic system; they influence in general the relative “weight” or “proportion” of each of the two subsystems. But we must not believe that plan and market will necessarily combine in a uniform, even manner within a given economic system. In this respect there may be essential differences, first by industries and then by type of decision problem. Let us now review the factors 4–6 which explain these deviations between industries and according to type of decision problem.

4. „Indivisibility,” increasing returns, standard versus fundamental decision. The nature of the decisions is different in the cases of a textile mill deciding to produce 100,000 or 105,000 meters of fabric next week and an electric company deciding whether to create a new hydroelectric plant. The first decision is largely controlled by the market. In making the latter decision, however, it is not customary to reason that “the price of electricity has risen, so let us build a new power plant” or that “the price of electricity has fallen, so there is no need for a new investment in power plants.” Instead, efforts are made to assess the future development and structure of the economy and to analyze the future demand for energy. This example contrasts two simple cases and shows that the two differ from each other in a whole series of characteristic features. These different features are surveyed in a tabular form in table 2-2.

5. Uncertainty. The more complex the decision problem and the longer the period which must be anticipated, the greater is the importance of the reliability of the forecast. The decisionmaker, as has been pointed out earlier, makes efforts to reduce uncertainty mainly by collecting information. National economic planning is a tool serving this end.

Factor 5 is related closely to the fourth one. The risk involved in the decision is great if it is connected with some large indivisible unit (in our former example: the building of a big power plant), and at the same time the decision must be taken with deficient information about the expectable consequences. The greater the risk, the more the decisionmaker will feel the necessity of collecting thorough and many-sided information.

6. Effects measurable and nonmeasurable with price. In every society there are inputs and results, favorable and unfavorable, which are not evaluated directly in terms of money and therefore have no price.

Western literature deals with this problem under two headings. One is “externalities,” which are contrasted to “internal” effects appearing in the calculations of the profit maximizing firm or of the household drawing up its budget (i.e., with all inputs and outputs whose effects are measurable with prices). The firm pays higher wages to its workers, if working conditions are polluted, but it pays nothing to the population if the factory pollutes the air and that air settles down on houses and gardens in the surrounding area.

The other heading is public goods. Here belong urbanization, water regulation, the preservation of
Table 2-2. Effect of the Type of Decision on the Combination of Market and Planning

<table>
<thead>
<tr>
<th>Item</th>
<th>Factors which favor the market subsystem</th>
<th>Factors which favor the planning subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of concentration</td>
<td>Little concentration, atomized market</td>
<td>High concentration</td>
</tr>
<tr>
<td>Character of decision</td>
<td>Standard decision (smaller modification in real variables against earlier situation)</td>
<td>Fundamental decision (major modification in real variables against earlier situation)</td>
</tr>
<tr>
<td>Indivisibility</td>
<td>The real variable controlled by the decision is continuous; there is no essential indivisibility</td>
<td>The real variable controlled by the decision is not continuous; there are essential indivisibilities</td>
</tr>
<tr>
<td>Character of the input-output function</td>
<td>There are no increasing returns</td>
<td>Increasing returns assert themselves</td>
</tr>
<tr>
<td>Time horizon</td>
<td>Decision can be taken with a short time horizon</td>
<td>Decision can be taken only on the basis of a longer time horizon</td>
</tr>
</tbody>
</table>

works of art, and so on. Nobody pays for enjoying the beauty of a city or for warding off a flood; yet everybody benefits from them.

One of the fundamental problems of the capitalist economy is the neglect of externalities and public goods. The extent of this neglect differs by countries; perhaps the most extreme examples can be found in the United States. Sociologists, economists, and politicians are much concerned about the grave troubles caused by pollution of rivers, lakes, and the air, by increasing noise, and by decreasing public safety. One way out of these difficulties is to internalize the externalities; firms should pay for effects which were not priced in the past. It is in the United States that one most frequently finds tolls on roads and bridges; it is in the United States that instead of general health insurance, the health service is based most on costs and there is no institutional care for old people. A "dollarization" of every human activity increases efficiency somewhat and forces economizing on resources; however, it leads to many anti-humanitarian, rigidly materialistic, and even outright irrational and wasteful phenomena.

In the socialist economies there is the opposite problem. Activities were not accounted for in money terms where it would have been expedient to do so. The effect was reduced to income incentives for working. Economists frequently have pointed out that the share of benefits above wages paid, "outside the envelope," is too great in comparison to the consumer goods available for purchase.

Although there is no delimitation that would be generally valid between the "internal" and the "external," between the spheres accounted for and not accounted for in prices, it is certain that there always will exist an external sphere. And with the growing wealth of society this sphere is bound to grow; this will become a historical tendency. It is possible (and in certain cases justified to some extent) to internalize some of the external effects, to account for them through prices, in terms of money. But in reality this can be done with most externalities only in forced manner. Their true regulator must be planning; the natural form of their description is information of nonprice character. One may argue whether tolls should be assessed on large highways. However, it is certain that one cannot assess a toll at the corner of each side-street. Therefore, the development of the road network must be planned as an integrated whole.

One of the advantages of a socialist economy is that the dominating role of social ownership relations allows increased possibilities for planning.

Let us sum up what has been said.

There exists no uniform combination of the control subsystems of the market and national economic planning independent of age and endowments. Their relative weights depend primarily on the political power and ownership relations. In addition, as regards the entirety of the system, the weights are greatly influenced by economic policy (growth rate and living standards) and by the general state of the market (pressure or suction). Within a given system planning is given a greater role in more highly concentrated industries, in fundamental decisions involving greater risks and relating to indivisible major units, and in the control of effects not measurable with prices.

A common characteristic of the factors listed hitherto is that they express circumstances objectively existing at a given historical moment. However, in addition there are also subjective factors at work,
operating through intentions of those controlling the economy. Today the role of planning is greater in the Netherlands than in Belgium, and this cannot be explained by differences in the situation of the two countries; much more important are the differing opinions of Dutch and Belgian politicians and economists. Similarly, if the role of the market in Hungary is different from that in Poland, this fact should be explained, not so much by objective differences between the two systems as by differences between the opinions formulated in Budapest and in Warsaw.

Comparison

Now that we have commented on the relation between market and planning, let us return to discussion of the GE (general equilibrium) school.

The partial market models of the GE school relying on the idea of "perfect competition" may be accepted as an approximate presentation of the interaction between supply, demand, and prices, valid for a special, narrow class of markets. This special, narrow class is characterized, among other things, by the following properties:

1. Both the demand side and the supply side consist of a great number of atomized organizations.
2. There exists no lasting tendency for the reproduction of market tensions, for constant disequilibrium, and for the preponderance of either side of the market.
3. There are no increasing returns. There are no large indivisible units. Production (and with it, sales) can adapt continuously to small changes in needs, and conversely.
4. Prices may develop freely, according to agreements between sellers and buyers.

For example, in many countries the well-known trade of pigs and corn belong to this narrow class. This is a well-known case of the "cobweb problem." Time series are available and these show the interaction of prices, supply, and demand, according to the basic ideas of the GE model.

The problem is, however, that GE economists want to squeeze the whole world into a theoretical construction which is suited only for the theoretical description of a narrow class of phenomena. This attempt is to no avail, as the model cannot accommodate the world. The partial market model is capable of describing and explaining the functioning of a definite kind of partial market. But its extension into general equilibrium theory is unacceptable.

In modern economic systems there are markets that fit the GE model and also others for which the above assumptions 1–4 are not characteristic. In addition, combined functioning of all partial markets raises further problems.

Modern economic systems are controlled by a complex control sphere, and the market is only one regulator of it. The GE school suggests that planning is a superfluous luxury, that an "optimal price system" could be found, which in itself is sufficient control.

The task of science is not to seek the "optimal" simple regulator of an unrealistic Walras-world but to describe, explain, and improve the complex control system of the real economy.

In discussions of earlier drafts of my book the following idea frequently was raised.

True, the theories of the GE school cannot be accepted as an accurate description or explanation of a real economy. But it should be accepted as a plan for designing a new world. Should the leaders of a country find themselves in a position in which they are able to develop the functional mechanism of their system themselves—as happened in the drafting of the system-the system should be formulated according to the models of the GE school.

It is true that the descriptive and the normative approach can be separated to a certain extent. The world can be changed. But not even the boldest change can entirely neglect the realistic possibilities. We cannot arrange the control system of an economy in a way that does not reckon with the real motivations of human activities, the finite limits of human intellect and capacity, the complexity and intricacy of the system, and so on. Therefore, before putting forward normative proposals, careful descriptive explanatory analysis is needed. Marx mobilized the men of science not only to explain the world but also to change it. His warning is topical today. But its opposite is no less topical; it is not enough to change the world—it must also be explained. What actually exists must be understood before we take a stand on what should be. For example, if we find that in the Soviet Union, the United States, Albania, Yugoslavia, and the Netherlands there simultaneously exist autonomous and higher control processes, price and nonprice information, market, and planning, we may assume safely that their simultaneous existence is an objective necessity. We therefore should not propose an arrangement that stresses one-sidedly only half of each pair of phenomena.

Before making proposals, we have to study very carefully everything that has developed in different economic systems. It is not that these are necessarily good and no changes should be made. The reason that such studies should be made is that in everything that comes about "in a natural way" there must be something that is necessary for the survival, functioning, and development of the system that called it into being. These elements of the system may be improved, but they cannot be neglected because obviously they have a task to perform.

In the language of mathematical decision theory,
the problem of "descriptive versus normative theory" may be formulated as follows:

The task of descriptive theory is to explore and describe the set of feasible systems; normative theory desires to select the best element of that set. The school, if considered normative, proposes to choose an "optimal system," but in searching for that system it reaches outside the set of feasible alternatives.

The better developed the market system, the more efficiently information can be provided. Even in an advanced country, however, production and consumption decisions that are made with reference to the future must necessarily be based on incomplete information as long as future prices are uncertain. For some commodities, a futures market may exist in which expectations of future supply and future demand determine the future price.

In a developing country, however, markets are frequently only localized, fragmented, and compartmentalized—thereby providing less information. Or a market may not yet have been established for a certain economic activity. An incomplete or rudimentary development of the market system means that the "organizational framework" of the economy is inadequate for the transmission of relevant economic information. The development of market institutions can accelerate what might be termed "the learning process of the economy" or a country's "social capacity" to incorporate change. And this can facilitate better utilization of the resources already at hand. As Myint observes:

"The standard theory of the allocation of resources implicitly based on the assumption of a fully developed organizational framework ignores the considerable potentialities for increasing the productivity and output from the existing resources in the unorganized sector without any dramatic technological innovations. The general sources of increasing productivity through the improvements in the organizational framework are to be found in a greater degree of specialization and exchange and in the development of a more effective network of communications linking up the numerous small economic units and transmitting to them the relevant economic information about the available resources, techniques of production and market demand conditions outside their own localities."

Allocative Function

To allocate scarce resources among competing ends is the prime function of a price system. In response to a rise in demand in one sector of the economy, price will rise, profits may temporarily increase, and resources will accordingly be attracted to the expanding sector away from another sector in which demand is stable or declining. The price rise provides an inducement for reallocation of resources to the expanding sector. Similarly, a decline in demand would act as a disincentive, and resources would be withdrawn from the declining sector. The commodity composition of production is thus determined by the market mechanism in conformity with changes in the pattern of demand. And the allocation of labor, land, and capital inputs will respond to the changes in demand for final outputs.

The more perfectly are markets organized—in the sense of being more competitive—the more readily will the allocation of resources come about through

the incentives and disincentives provided by the market-price system. Under perfectly competitive conditions, resources would be reallocated from low-productivity activities where prices are low to high-productivity activities where prices are high. An optimal allocation of resources would fulfill the properties of efficiency: output would be maximized when the use of an additional unit of a factor of production of equal quality contributed the same increment to total output (value of marginal product) in each possible use and received the same reward in each activity.

Market imperfections, however, will impede the smoothness and flexibility of the allocative process. If, for institutional reasons or because of the monopoly power of a producer, prices are inflexible, then the full power of the allocative function is reduced. Similarly, if economic agents lack the knowledge of market conditions and opportunities, the allocative function cannot operate. Resource immobility can also be caused by sociocultural and political conditions.

A less developed economy is often portrayed as a dual economy, with a modern, organized sector and a traditional, unorganized, or informal, sector. In many developing countries this takes the form of a modern commercialized industrial sector that has developed alongside a traditional subsistence agricultural sector. In such an economy, market imperfections tend to be pervasive, and an efficient allocation of resources is inhibited.

Dualism will affect resource allocation across the entire spectrum of markets in a developing economy. The intersectoral allocation of resources between agriculture and manufacturing is examined in Selection 2.3.

2.3 Distortions of Agricultural Incentives

Most developing countries for the last quarter century have had policies designed to lower the prices of food and other agricultural goods and to increase the prices of manufactured goods. Trade and foreign exchange practices have been major instruments of these policies, along with tax, direct price, and other market-control measures. The conventional wisdom supporting this twisting of the terms of trade against agriculture has had four main pillars, based on the assumptions:

1. that aggregate agricultural production is not very responsive to price changes;
2. that the chief beneficiaries of higher prices would be the larger size farmers;
3. that higher food and other agriculture-related prices such as clothing would most adversely affect low-income consumers; and
4. that manufacturing provides a more rapid means of growth, and that achieving that growth depends upon large transfers of income (profits) and foreign exchange from agriculture to manufacturing. Thus, policies that depress agricultural prices and increase manufacturing prices will result both in more rapid economic growth and in a more equal distribution of income.

Lagging agricultural production and overall economic growth in many developing countries that have followed such policies, however, have led to increasing concern about measures that reduce farm incomes and incentives. Argentina, Egypt, Kenya, the Ivory Coast, Pakistan, Peru, the Philippines, Thailand, and Uruguay are among the countries that have acted to increase significantly agricultural in relation to nonagricultural prices in the past several years. Greater emphasis in the development literature on employment-led and rural-development strategies, and on strategies to meet basic needs and to benefit persons in the lowest 40 percent of the income distribution, has also provided new intellectual support for price policies more favorable to agriculture, including lesserened subsidies for capital goods, less overvalued exchange rates and protection for industry, and increased production of foodstuffs and other wage goods.


In contrast to those of the developing countries, price policies in the industrialized countries during this period have been distorted in favor of, rather than against, the farmer, and the typical problem has been overproduction rather than underproduction of agricultural commodities. Moreover, as pointed out by D. Gale Johnson, greater agricultural productivity per hectare in industrialized than in developing countries is a phenomenon that has occurred largely since the 1930s, and within the period of differential price policies. In the period 1934–38, grain yields averaged 1.15 tons per hectare in industrial countries, and 1.14 tons (i.e., the same level) in developing countries. By 1975, however, grain yields in industrial countries were more than double those in developing countries, 3.0 tons versus 1.4 tons per hectare. This is not to argue, of course, that the more rapid growth in agricultural yields and production in the industrialized countries is due solely to differences in price policies. A plausible hypothesis is that both the more favorable farm prices and the more rapid growth of farm yields in industrial countries reflect efforts to support farm incomes and increase agricultural productivity, while developing countries have generally been much more concerned with increasing industrial incentives and production, and urban incomes.

The focus here is not on the factors that explain the differences in growth rates between developed and developing countries, however, but on the effects of recent pricing policies in today's developing countries. My hypothesis is that agricultural production, income distribution, and economic growth would all benefit from reduction or elimination of distortions that reduce agriculture's domestic terms of trade. The relation of agriculture's domestic terms of trade to income distribution and economic growth particularly are problems whose complexities outrun our capacity to measure by formal models, but they are also problems that are too important to neglect.

Several general problems of terminology or analysis should be briefly mentioned. First, references to "free-market" prices are to those that would exist in the absence of specific controls and policies that now distort the relation of agricultural to nonagricultural prices, including those relationships that magnify differences between domestic and world-market prices. This is not to imply that world-market prices are "free," or that policy should aim to equate domestic and world-market prices or price relatives at all times. Neither is it to imply that shadow free-market prices can be known with exactitude, or to ignore that we are dealing with neither a first- nor even a second-best world. Rather, the comparison with probable prices "in the absence of controls" reflects only a judgment about approximate price relationships. Second, references to changes in prices are to changes from existing distorted price relations, not from the traditional equilibrium-price starting point of economic theory. Usually the discussion concerns what would happen if the distortion between existing prices and free-market or "equilibrium" (using that term loosely) prices were reduced. Third, prices are an incomplete measure of incentives. If agriculture's relative productivity is increasing, it is possible for agricultural incomes and incentives to be growing at the same time that agricultural prices are declining.

Fourth, a fundamental distinction must be made between measures such as price controls, taxes, and subsidies that artificially lower food and agricultural prices and measures such as on-farm investment, technological advances, and rural infrastructure development (e.g., roads, electricity, and water) that lower prices by lowering the real costs of production. Reductions in real resource costs of production (i.e., increases in input yields) may benefit urban dwellers, farmers, and rural laborers alike.

**Pricing Policies and Agricultural Production**

**EFFICIENT USE OF RESOURCES**

Higher prices may stimulate agricultural production by (1) causing producers to move closer to their production-possibility frontier by better use of resources, (2) encouraging use of more labor and other variable resource inputs to reach higher production-function and output levels, or (3) inducing investment and the discovery and adoption of new agricultural technologies that result in new, lower-cost production functions.

The usual empirical basis of arguments that agricultural production is not very responsive to price changes rests upon multiple-correlation analysis of the acreage or output of a particular crop in relation to the relative prices of that crop versus others, or to several presumably independent variables, such as output price and use of fertilizer, labor, and water. The price elasticity of aggregate agricultural production is less than for individual crops, of course, since substitution between crops may account for an important part of the response of a single crop to changes in its relative prices. This multiple-correlation methodology can be faulted on several grounds, including the lack of independence between prices on the one hand and the level of use of fertilizer and other inputs on the other. The use of acreage rather than output in most such studies underestimates the price elasticity of production by not taking account of changes in yields, which have been accounting for an increasingly large part (now more than half) of the annual growth in world food-grain production. Also, it is doubtful that yearly fluctuations in prices are an adequate proxy for changes in the income expecta-
tions that determine rates of investment in agriculture and its supporting infrastructure, labor inputs, and the adoption of new techniques. These decisions appear to depend more upon whether expected profitability is above or below a "threshold" level of acceptability, and may be little affected by yearly price fluctuations unless these cause longer-run expectations of profitability to change. Moreover, incentives are a function of net income and not just prices, and therefore a new technology may importantly change the incentive effect of a given set of prices, as happened during the Green Revolution in wheat and rice.

Price incentives may cause farmers to use improved seeds, along with more fertilizer, pesticides, and other purchased inputs, to adopt improved cultural practices, and to apply more family or hired labor. All of these ways of increasing the efficiency of resource use may occur at once, for example, if a small farmer decides to reduce or give up his off-farm employment—or a son does not migrate to the city—in order to adopt the more labor-intensive techniques required to reap the benefits of new varieties. Generalities about most small farmers as "subsistence" farmers unaffected by agricultural prices are at best misleading.

Timmer and Falcon have found a close rank correlation between unhulled rice (paddy) prices and rice yields among Asian countries. For 1970, they found that rice yields varied from 5.64 and 4.55 metric tons per hectare in Japan and Korea to 2.1 to 1.7 tons per hectare in Indonesia, Thailand, the Philippines, and Burma. At the same time, the ratio of the price of a kilogram of rice to the price of a kilogram of fertilizer nutrients showed a similar wide variation, from 1.43 and 0.96 in Japan and Korea to 0.4 to 0.1 in Indonesia, Thailand, the Philippines, and Burma. This pattern of intercountry relationships between prices and yields cannot be explained as short-term elasticity response, but may well be indicative of the long-run responsiveness of production to incentives. While the price data used in this study were for only one year, the differences in national prices are very substantial and reflect long-standing differences in price policies among these countries. It is noteworthy that the three areas with the highest yields—Japan, Korea, and Taiwan—have some of the poorest soils among the nine places studied. The high prices to farmers there appear necessary to cover the costs of achieving these yields under the given climatic, soil, and other conditions, even though the decisions to pay such high prices may have been primarily political. It is interesting to speculate to what levels rice yields and production would fall in Japan if farmers there received the same price as Thai farmers, or conversely, the levels to which rice yields and production would climb in Thailand within a few years if Thai farmers faced the same rice and fertilizer prices as Japanese farmers.

**SHIFTS IN THE PRODUCTION FUNCTION**

The most important long-run effects of price incentives on production may be through price-induced shifts in the production function, rather than through greater efficiency of resource use with existing production functions. These long-run effects depend upon the extent to which the incentive structure has an important effect upon technological change (both on research and on adoption of new techniques by farmers), on public and private investment related to agriculture, and on institutional change affecting agricultural output (land reform, extension services, marketing, and distribution facilities).

Clear links have been demonstrated between pricing policy and both the discovery and nature of new technological discoveries, and the adoption of these techniques by farmers. The relationship between relative prices of land and labor in Japan and in the United States, and the very different directions of technological change in these two countries in the last century, have been well documented. The late-1973 jump in oil prices has also presented ample evidence of the link between prices and research. Much of the agricultural research initiated since then is "energy saving," with heavy emphasis on reducing fertilizer requirements, for example, through nitrogen fixation, seed treatment, and placing a small amount of fertilizer close to the seed or seedling roots.

The link between price incentives and private investment in chemical inputs, labor, land leveling, irrigation, and other measures to increase farm output is almost axiomatic. Studies of diffusion of new techniques and of new investment in developing countries indicate that where profitability of adopting a new technique or investing, say, in a tube well is very high, the new techniques will be rapidly adopted and the new investment quickly made. At the margin at least, there are wealthy individuals who make choices between agricultural and nonfarm investments. A large proportion of upper- and even middle-class individuals in South Asia reside in cities and have a profession, a business, or a government job, but own substantial farmland operated by a hired manager or tenant (sometimes a relative). For these individuals, as for the small farmer deciding how to divide his time between his own farm and other employment, the choice between farm and nonfarm investment (and employment) is a recurring one. Higher returns to farming certainly attract more labor and more investment (ceteris paribus) into farming. Moreover, the higher returns generate more income and saving and, therefore, the ability to finance more investment in agriculture. Nearly all studies show
high marginal saving rates among even poor farmers. Thus, an important part of the additional income accruing to farmers through higher prices is likely to result in greater savings and investment. In countries where farm yields are low and returns to additional investment in agriculture are high (e.g., where fertilizer use is well below optimal levels because farmers lack financial resources, as is true in much of South Asia and other food-deficit low-income countries), an important part of that saving is likely to be invested in farming.

Public investment and institutional changes may also be importantly affected by prices, but research on these topics is in its infancy. Evidence has been presented that public investment in irrigation in the Philippines has been closely correlated with the import price of rice. There has certainly been a tendency for some officials to think of nonagricultural development as somehow more important than progress in agriculture. Higher prices for agricultural products will make at least the non-shadow-priced value of agricultural output greater, and the nominal or financial rate of return to agricultural projects more attractive. The more attractive financial return may induce highly productive institutional development, such as more effective agricultural research and extension systems, and input distribution and output-marketing systems.

Rationing Function

If prices could not rise when demand rises or supply falls, the excess demand could be suppressed only by direct allocation of the scarce supply through rationing the limited supply among the potential buyers. As previously noted in Selection 1.5, if the government imposes a price ceiling on a product, there will be a shortage in the market, and this may commonly lead to rationing.

When, however, the price of a scarce commodity or factor does rise, the price increase in itself performs the rationing function. Only buyers who place the higher value on the commodity or factor will exercise their demand. Other buyers who have a lower valuation of the commodity or factor will leave the market. While the price rise does reduce the quantity demanded to the smaller available supply, it may do so by eliminating buyers who can no longer afford to pay the higher price, as well as some buyers who may now prefer a lower-priced substitute. When low-income buyers are squeezed out of the market, the government may believe that to provide an equitable distribution of the commodity, a direct form of rationing may be preferable to rationing via a price rise. But to the extent that price is not allowed to reflect the relative scarcity value of the commodity, some governmental intervention in the market will be necessary to deal with the market shortage of the commodity. The costs of this intervention must be recognized: administrative costs, possible costs of corruption, and the costs of inefficient resource allocation. These costs must be weighed against the objective of equity.

The phenomenon of a market shortage is common in developing countries. Recourse to price ceilings, especially for foodstuffs and primary products, has already been noted in Selection 2.3. In another market, the maintenance of an overvalued foreign exchange rate has also caused the government to ration scarce foreign exchange by exchange licenses or imposition of tariffs and quantitative restrictions on imports. Instead of allowing the domestic currency to depreciate in terms of foreign exchange, the government resorts to direct specific controls to reduce the quantity demanded of foreign exchange. Again, this type of price distortion operates against the efficient allocation of resources as between import-substitute commodities and exports of nontradable goods.

When the government imposes a tax on a commodity, the price will rise and the quantity demanded will decrease. The tax facilitates the rationing function. If the tax is initially levied on the seller, the amount of the tax that is shifted forward to the buyer will depend on conditions of elasticity in demand and supply
in the particular market. If demand is perfectly inelastic, the incidence of the entire tax can fall on the buyer. If demand is perfectly elastic, the seller is unable to raise price above the market price. But in between these extreme cases, both the seller and buyer will absorb some of the tax. A sales tax on a commodity will raise the price most and reduce the quantity bought least when the demand is more inelastic than supply. When the supply curve is more inelastic, the price will rise less to the consumer, and more of the tax will be absorbed by the producer.

The rationing function of prices is highlighted by the interesting problems of what price to charge for an essential public service (Selection 2.4), and how to ration the use of congested roads, as considered in Selection 2.5, and the use of ports as in Selection 2.6.

2.4 Pricing of Public Urban Services

Why not provide essential public services such as water free of charge and finance their costs out of tax revenues and government borrowing?

Perhaps the first reason that comes to mind is that the amount of water that cities in developing countries could afford to supply would fall considerably short of what people want or need. Indeed, in many cities in developing countries, a large proportion of the population has very little or no access to potable water. Hence, some method has to be devised for distributing the limited quantity of water among the population. It calls for some sort of rationing system. If the government does nothing, as often happens, the water will go to the people who get to the taps soonest with the biggest containers. But this approach is likely to be disorderly and unfair.

One method of distributing the water in a reasonably orderly manner is to charge a price for it. The water would then go to those who were willing and able to pay the price, and the quantity of water obtained by a user would depend on how much he was prepared to spend for it. What would happen if the demand for water exceeded the supply at a particular price? The price would simply be raised to discourage demand until it was in line with the available supply. Pricing would be more efficient than a rationing system that tried to distribute the water according to need because it would require much less administration than a rationing system and hence would be less costly to the public authorities.

The distribution of water through pricing would also be more efficient than rationing from a purely economic standpoint because the water would yield more economic value. But how does the economist judge how much value is produced by a resource such as water? The simple answer is that it depends on the price that users are willing to pay for it. The higher the price they are willing to pay for water, the more economic value they derive from it. This is one of the central propositions of economic theory. But what about the people who are unable to pay the price of satisfying their minimum water requirements, or, if they do pay it, will be unable to satisfy their minimum needs for other goods and services? The answer given by price theory is that pricing would generally be a more efficient way of distributing water than rationing because the costs of distribution would be lower and the economic value of the water would be greater. But it would not necessarily be a fairer way of distributing it. There is a difference between economic efficiency and distributive justice. They are both important objectives, but they often conflict with each other.

The notion that pricing is a more efficient method of distributing water than providing it free of charge or rationing rests on the assumption that pricing is more likely to distribute water to people who will put it to the "best" uses so that there would be less waste. What is meant by "waste" in economic terms and how would we identify it? It is easy enough to identify the most obvious cases of waste, such as leaving a water tap running without using the water. But what about the use of water for washing cars? What criterion would we apply in trying to decide whether this would be a wasteful use of water? The economist's answer is that waste has to be judged in relative and not in absolute terms. Hence, washing cars would be a wasteful use of water if this water could be used in more valuable ways elsewhere. How could we tell whether the water has more valuable alternative uses? By the price that users were willing to pay for it, so that if farmers were prepared to pay a higher price for water for irrigation than urban users would pay for washing cars, then the water would yield more economic value if it were used for irrigation. But what does this have to do with the relations between pricing and waste? The answer is that if the users of water pay nothing for it, they have very little in-

centive for putting it to its most valuable uses. By using less water, they save nothing. Why take the trouble to turn off the tap if it costs you nothing to leave it on? Thus, in the case of public goods and services for which there is ample opportunity for wasteful use, pricing tends to reduce waste. Indeed, the scope for waste is one of the important criteria for deciding whether to charge a price for public services or to provide them free of charge. Thus, one of the reasons national defense, police, and fire-fighting services are provided free of charge is because there is very little scope for the individual citizen or family to use these services wastefully. But when it comes to providing free public health services, for example, the case is less clear.

Since pricing is more likely to produce a more orderly and efficient distribution and less wasteful use of water than providing it free of charge or rationing it, what criteria should the public authorities apply to decide how much water to supply? One answer is that they should supply as much as is needed. How would need be measured? If it included more than what is absolutely essential, it would be more than most cities in developing countries could afford to supply. One approach would be to supply the minimum amount required per person if the authorities can afford to do so. Such rationing would require a fairly costly administrative and technical apparatus. It might supply some families with less and others more than they need. Efforts to correct such inequities would produce a very complicated distribution system. But what about uses of water over and above the minimum requirements of households? What guidelines should the authorities use to decide how much water to supply? With rationing, the guidelines would have to be quite arbitrary. Would pricing provide a less arbitrary guideline for deciding how much water to supply? It would in the sense that users could indicate how much water they were willing to use at a particular price. These quantities would be less than they would consume or “need” if the water were provided free of charge. But it might still be more than the authorities could afford to supply. By charging a higher price, this gap would be reduced. What the “right” price and supply should be will be discussed below. The point here is that pricing is a device for measuring users’ “need” for water and thereby helps the authorities decide how much to supply.

If public goods or services are supplied free of charge, the costs of producing them have to be financed out of tax revenues or public borrowing. But charging a price will produce revenues that would reduce the reliance on budgetary financing. In fact, the supply of urban services in most developing countries is heavily dependent on government financing. This is an unreliable source of financing and often accounts for the relative scarcity and poor quality of urban services. Thus, greater reliance on self-financing to pay the costs of supplying urban services will often produce more and better services.

Budgetary financing of urban services may be inequitable in its effects on income distribution. If the costs of supplying urban services are financed by imposing or raising sales taxes, the taxes often hit the poor harder than the rich and hence are likely to be more inequitable than pricing. Also, if the financing is done by diverting public funds from other programs, the funds are often diverted from economic and social programs whose benefits accrue mostly to the lower-income groups. Thus, while the provision of free water, for example, is often believed to be more equitable than charging for it, the actual effect may be the opposite.

The above review of the major consequences of providing urban goods and services free of charge and financing their costs out of budgetary funds has identified the following main functions of pricing:

1. It provides an orderly and economically efficient, though perhaps unfair, method for distributing a limited supply of goods and services.
2. It provides an incentive for reducing wasteful uses of the goods and services.
3. It provides a guideline for the suppliers of the goods and services to help them decide how much to supply.
4. The revenues produced by pricing lighten the burden on government budgets and often lead to a more equitable method of financing the costs of public services.

Charging Too High a Price: Evidence and Consequences

What evidence would suggest that the price being charged for water is too high? One common-sense possibility would be that people were using less water than the water company was able to supply. The company’s capacity to supply additional water depends upon whether it can finance the costs of doing so. In most urban water systems, most of the costs of supplying water are capital or fixed costs in the form of reservoirs, pumps, pipes, and treatment plants. This means that as long as the supply of untreated water is adequate and the treatment and delivery equipment is not operating at full capacity, the costs of supplying additional quantities of treated water are small. These incremental costs are called marginal costs by economists. If a water company can supply an additional quantity of water at marginal costs that are below the price that users would be willing to pay, then the company would be regarded as having excess or unutilized capacity. If this price is less than the price being charged, it is reasonable to conclude that the price of water is responsible for the excess
capacity because a lower price would increase the amount of water demanded and supplied and thereby lead to a utilization of the excess capacity.

From an economic standpoint, excess capacity is a misuse of resources because the capital equipment that is not being used to supply water is not producing any economic value. Moreover, the additional resources required to supply the additional water, i.e. the marginal costs, might be less than the price that users would be willing to pay for the water. This price is assumed to be a measure of the economic value that users and the economy would derive from these resources. The marginal costs are assumed to be a measure of the economic value that these resources would yield in alternative uses. The economy would therefore obtain more value from these resources if they were used for supplying additional water than from alternative uses. Thus, charging too high a price for water can produce excess capacity that would result in an uneconomic use of resources.

Another possible consequence of an excessive price for water is that it may prevent low-income groups from getting a minimum supply of water and lead them to dig their own wells or to use contaminated water that can produce serious health hazards. While some low-income families may be able to buy their minimum water requirements, a high price may absorb a disproportionately large share of their incomes. As for industrial uses of water, where the cost of water accounts for a significant part of the total costs of production, a high price for urban water could discourage certain types of industrial production. But these water-intensive industries are rather rare.

Charging Too Low a Price: Evidence and Consequences

We have seen that one of the possible consequences of charging too high a price for water is that users will buy less water than the system is willing and able to supply. Charging too low a price could have the opposite effect, leading to excess demand. If the water system were operating at its full capacity, the excess demand could be eliminated by raising the price and/or expanding capacity. The effectiveness of the former would depend on how responsive the water users would be to an increase in price, i.e. the elasticity of demand. Whether it would be worthwhile from an economic standpoint to expand capacity would depend on how much economic value would be derived from the resources that would be required to expand capacity compared with the economic value that these resources would yield in alternative uses. According to traditional economic theory, the latter would be measured by the cost of these resources. The amount paid by the users of the water produced by the additional capacity would be the measure of the economic value that these resources would produce if they were used to supply water.

If it were not feasible to eliminate excess demand by raising the price and/or expanding capacity, other alternatives would be to ration or allow nature to take its course. We have seen that both alternatives present major problems.

Another possible consequence of charging too low a price for a public service is that the amount of revenue produced from the sale of the service does not cover the operating and maintenance costs of the enterprise, so that unless the deficit can be financed by government budgets or some other source, the quality of the service is likely to deteriorate. Moreover, if the enterprise has to rely entirely on outside sources for financing the expansion of its capacity, it is likely to encounter serious difficulties.

What Is the Right Price?

We have seen that the provision of certain urban services free of charge can produce chaos, waste, and injustice. We have also seen that charging the "wrong" price for these services can also produce problems. Having established the desirability of pricing these services, what then is the "right" price or pricing policy? A simple but imprecise answer is that it is the price or policy that leads to the production of the right amount of the service and its distribution to the right users in the right amounts and with the right quality. But this is nothing more than a restatement of the functions of pricing. It is also a statement of the criteria for assessing the merits of pricing policy for urban services. While most public authorities would be willing to accept these criteria, there are two very big problems in applying them. One is that they are not stated rigorously enough to be applied with acceptable precision. The other is that the functions of pricing are often in conflict with each other. For example, a pricing policy may lead to the production of the right amount of the service, but much of it may go to the wrong users. Hence, the decisionmakers face the task of reconciling these conflicting consequences. This is often more a matter of judgment than rigorous analysis because the analysis is partial in its coverage, the causal relationships are uncertain and imprecise, and the conflicting consequences are not measurable at all or not in the same units, so that their net effect cannot be determined with precision. These problems are inherent in many issues of public policy and help explain the limitations imposed on the use of rigorous analysis in decisionmaking.

One way of avoiding the problem of reconciling the conflicting consequences of pricing policy is to concentrate on only one of its objectives and to pay relatively little attention to the others. For example,
a pricing policy that aims to produce an amount of revenue that will cover the total costs of supplying the service and also yield a surplus to cover part of the costs of expanding capacity is likely to pay relatively little attention to the income distribution consequences of the policy. One approach is to deal with these consequences but through methods other than pricing. Another approach is to reconcile conflicting objectives and consequences through differential pricing practices and other devices to achieve acceptable compromises.

2.5 Congestion Pricing—The Example of Singapore

The motor vehicle and its use as a private means of transport have profoundly affected the style and character of life in many countries by expanding the range of opportunities for work, recreation, and social activity. There are, however, adverse effects that cannot be ignored. The increasing use of motor vehicles has contributed to environmental pollution, to balance-of-payments deficits for oil-importing countries, to a deteriorating quality of life in city centers, and to congestion on limited urban road space. In recent years, it has become increasingly obvious that the economic, social, and environmental costs of urban road construction programs are unacceptably high, especially for cities in the developing world where resources are scarce. Moreover, even the largest road construction programs have failed to solve the problem of congestion, as demands for road space continue to grow faster than space can be expanded.

Since increasing the supply has failed to balance demand and supply, greater consideration has recently been given to the possibilities of containing the demand by methods generally referred to in the transport planning profession as traffic limitation or restraint. These methods include giving preferential treatment to high-occupancy vehicles, reserving streets for pedestrians and buses, imposing delays on drivers of private cars, and road pricing. Attempts to encourage people to ride public transport or to form car pools have met with limited success, as have efforts to make car travel less convenient by imposing detours and delays. Thus, transport planners have turned to the more direct approach of making people pay more for journeys in vehicles that use scarce street space inefficiently, specifically cars with only one or two people in them.

The use of road pricing to restrain inefficient use of motor vehicles, to conserve scarce capital resources and fuel, and to avoid adverse environmental impacts may also be viewed in purely economic terms. In other sectors of the economy, demand and supply are balanced by market-determined prices. This mechanism does not function adequately in the transport sector because motorists generally are not required to pay the full social cost of any particular trip, that is, the costs imposed on others in terms of delay, air pollution, noise, and so on. The result is that, for many car trips, the cost to the traveler is less than the benefit he gets from the trip, while the total cost to society exceeds the benefit. When this is so, trips take place that—from the social viewpoint—should not be made and thus contribute to congestion. Road pricing is a way of making the motorist recognize the entire social cost of any trip by requiring him to pay the difference between that cost and the cost he would otherwise perceive. This should inhibit him from making trips for which the social costs exceed the benefits.

Car usage interferes not only with other cars. In addition, free access to road capacity leads to a level of car usage that has adverse effects on public transport. A vicious circle develops in which congestion resulting from the increased use of cars leads to poorer public transport service and, hence, reduced ridership. People give up using the unreliable bus service and drive cars instead. This increases the number of cars on the road, further degrading the public transport service and contributing to deficits for public transport operators. In general, they respond by increasing fares, or cutting service, or both, which exacerbates the problem. Increased public transport efficiency could result from reducing the amount of private car usage and promoting the use of public transport.

Finally, it should be noted that the present system of underpricing road use generally favors those who can afford to purchase cars and has an adverse effect on poorer people who are obliged to use a public transport system with declining service levels.

Despite widely expressed interest in urban road pricing by transport and urban management experts, a variety of potential problems—political, administrative, and enforcement—have generally deterred public officials from adopting this approach to urban transport problems. At the time of this writing, Singapore is the only city in the world that has implemented a road-pricing scheme. The scheme was introduced in June 1975. Its essence is that a special supplementary license must be purchased and displayed on any car that is driven into a designated restricted zone during the morning commuting hours. The license scheme was complemented by increased downtown parking rates, and measures were taken
to provide improved bus service, including park-and-ride arrangements.

Implementation of the scheme went very smoothly. Contributing to this success were a well-conducted, year-long public information campaign, excellent planning and management of details, and quick recognition and correction of the few problems that did arise. Also helpful was the disposition of Singaporeans to believe that their government was acting in the general social interest and to abide by the rules imposed.

Unlike most government actions to improve transport service, Singapore's area license scheme is profitable. Initially, revenues (equivalent to nearly US$200,000) were nine times current costs, and the rate of return on the investment (equivalent to US$2.73 million) was about 77 percent a year. In 1976, after the fee was increased, the rate of return approached 95 percent.

Achievement of the Objectives of the Area License Scheme

At the most general level, the objective of Singapore's transport planners was to contain the growth of traffic congestion to prevent undesirable effects on road users and on the environment within the city. According to the publicity booklet distributed by the Road Transport Action Committee: "Daily traffic congestion results in delay and frustration to motorists, bus commuters, goods and emergency vehicles, and poses danger to pedestrians and other road users. It also causes deterioration of the environment through noise, air pollution, and visual blight."

Before the introduction of the area license scheme, off-peak traffic volumes were about 25 percent lower than peak volumes and were considered acceptable. On this basis, the transport planners set a target of a 25 to 30 percent reduction in peak-hour traffic volumes entering the restricted zone. This corresponded to a 50 percent reduction in the number of cars entering the restricted zone. The area license scheme achieved considerable success in reducing congestion in the central area, largely by inducing the expected shifts toward public transport and car pools, by spreading the peak, and by diverting through traffic.

In fact, it was more successful than planners and observers had expected, achieving a reduction in the number of cars of more than 70 percent and a reduction in total traffic volumes of more than 40 percent. The achievement of a reduction in excess of the target has resulted in a high degree of under-utilization of roads inside the zone, and observers have concluded that the price was set too high. The transport planners in Singapore, however, have not reduced the fee, believing that achieving their long-run objective of modifying attitudes is more important than achieving short-run economic efficiency. In short, having got people out of their cars, they have no intention of inducing them back again in the name of economic efficiency.

The target reduction in traffic volumes was intended to reduce "delay and frustration" and avoid continued "environmental deterioration." These benefits were not articulated in quantitative terms by the planners of the area license scheme, but, rather, in such terms as being "freed of the burden and frustration of crawling through traffic jams in the city." At this subjective level, motorists in the public opinion survey reported large improvements in the level of congestion, traffic speeds, and travel times in the restricted zone during the restricted hours. Bus riders, taxi riders, and motorcyclists also reported improvements in these categories. Thus transport users themselves perceive traffic conditions to have improved and the frustration associated with travel to the restricted zone to have been diminished.

At a more quantitative level, the speed study revealed a 20 percent increase in mean car speeds in the restricted zone during the restricted hours. This increase, however, was not reflected in the travel time changes derived from the household survey data. According to that source, car drivers overall lost small amounts of time, and bus riders gained only small amounts. Thus, it may not be concluded that the area license scheme led to significant reductions in delay. Nevertheless, the future time savings associated with keeping congestion levels constant rather than allowing them to get worse (for example, under the traditional "minimum intervention" assumption of economic analyses) are likely to be extremely large.

Air pollution levels in the restricted zone have been reduced. The likelihood that a pedestrian will have to take evasive action to avoid being hit by a car has been reduced. According to the public opinion survey, people perceive noise to have been reduced, safety when crossing roads to have been increased, general shopping conditions in the central business district to have improved, and the effect on Singapore as a city to be favorable. It may be concluded, therefore, that the area license scheme has had a positive effect on the environment.

It should be remembered that traffic congestion in central Singapore was by no means extreme before the introduction of the area license scheme, and the scheme was introduced to prevent the situation from becoming severe. To achieve the long-run goals, the planners perceived that, in addition to explaining the rationale for more widespread use of public transport and other high-occupancy vehicles, they would have to induce motorists to review and fundamentally change their attitudes toward the ownership and use of cars. This revision of motorists' attitudes and,
hence, behavior, was expected to reduce the problems caused by congestion and at the same time create an environment in which public transport services could be improved.

It is clear that the area license scheme forced motorists to modify their behavior, at least in the short run. While the high level of taxation on cars has stabilized car ownership levels, the area license scheme has reduced congestion in the central area, largely by inducing shifts toward public transport and car pools, together with shifts to earlier times and routes that avoid the restricted zone. Whether these are simply short-term behavior modifications or whether they really represent fundamental changes in the attitudes of motorists cannot be determined at this point. It seems likely, however, that the continued use of such measures will result in a more widespread acceptance (rather than mere tolerance) of public transport, car pooling, and other alternatives in the long run.

The benefits of meeting these general objectives cannot be valued in money terms, but the creation of a breathing space and a significant probability that Singapore can be prevented from becoming intolerably congested is clearly an important and valuable achievement.

Problems

While the area license scheme, by and large, achieved the primary objectives of its designers, some of its effects did not turn out as expected; some problems arose and had to be dealt with.

LEVEL OF THE LICENSE FEE

With no precedent to guide them, the officials had to set the fee on the basis of an "educated guess." That they set it within the right order of magnitude showed very good judgment. Nevertheless, in terms of the target specified they set it somewhat too high. That was probably better than setting it too low, but it did pose the question of whether to reduce it or leave it alone. Choice of the latter alternative was justified in terms of conditioning people's attitudes so as to postpone future congestion problems.

THE EVENING PEAK

Singapore officials, the World Bank research team, and consultants all believed that if morning peak traffic were reduced by the targeted amount, evening peak traffic would also drop considerably. All were surprised when that did not occur. There were discussions of the possibility of imposing restraints of some kind in the evening, either on inbound or on outbound traffic. However, there seemed to be no way to do that without unfavorably affecting business—for example, in the restaurant and entertainment sectors as well as retail trade. No good solution to this problem was found, and no restraints have yet been imposed in the evening.

PARK AND RIDE

The designers of the scheme believed that former car drivers looking for a different mode of transport to avoid the license fee would be attracted by the possibility of driving to the outskirts of the central business district, parking, and taking a shuttle bus downtown, especially if they could be assured of a seat on the shuttle bus, a short wait, and a reasonably quick trip. On this basis, they spent S$6 million (US$2.5 million) preparing fringe car parks with capacity for 10,000 cars. This was the major expenditure connected with the scheme. They also awarded franchises for the operation of shuttle buses, which were restricted to operating on specified routes from the fringe car parks with frequent service, limited stops, and no standing passengers. With the projected demand, this would have been a profitable operation.

In the event, these park-and-ride arrangements were not attractive to Singapore drivers. About 6 percent of the parking space was used, and the shuttle buses were so underutilized that the operators were in immediate financial difficulties. The government acted quickly to revise the shuttle bus operations, making them essentially a supplement to the regular bus service on profitable routes where more capacity was needed. In time they found new uses for much of the car park space.

2.6 Congestion Levies and Surcharges for Ports

In many developing countries the high demand for very limited port services causes congestion. This suggests that there are good reasons to earn high profits in the ports by raising congestion charges. If the demand for a service is very high, in relation to normal conditions, the price of that service should be increased so that users have the incentive to economize on the scarce service. Port authorities rarely, if ever, respond in this way. Normally, they keep prices low. Thus large queues of vessels form in the

Selection 2.6 reprinted from Esra Bennathan and A. A. Walters, Port Pricing and Investment Policy for Developing Countries (New York: Oxford University Press, 1979), pp. 7–8, and 81.
roads, and shipowners are quick to add their queuing cost to the freight rates. The domestic producer and consumer pay for the real cost caused by scarce port capacity and the fact that the port authority does not raise its tariffs.

The main issue is, however, a positive one: what would be the consequences of the port's raising a congestion levy? And more germane still, would the residents of the country be better off than under the low port price and queuing regime? Generally, the price of the overall shipping service (freight rates plus port charge) will rise above the value that would apply if the port did not charge a congestion levy. The increase in price, however, is likely to be considerably less than the congestion levy. If the demand for port services is inelastic, the port will increase its revenue—even when shipping is a monopolized service—at the expense of domestic traders. Congestion charges will give the authorities a source of finance to extend port facilities. And they may also increase the throughput of the port considerably if the congestion charges are levied only on those particular facilities (such as transit sheds) that are congested.

From this survey of congestion, it is concluded that there is a powerful case for the ports to charge congestion levies. However, the case is based on our general understanding of economic forces and on a close study of the nature of the shipping business and its interaction with the domestic economy; we cannot point to the experience of any port which has introduced such charges. Until such crucial experiments have been made, the case for such charges cannot be substantiated.

Advantages of Congestion Pricing

One main advantage of congestion pricing is that the port authority appropriates the surplus caused by the demand for its facilities. It is not uncommon for ports, particularly in developing countries, to be congested and simultaneously to show large losses even on operating account; levying congestion charges would avoid this undesirable state of affairs. Second, the levies would provide funds to expand port services. This automatic financing arrangement would make it more likely that the port, like the tub, could stand on its own bottom. If, however, the port is subsidized, then such funds must be raised by taxation; this not only incurs an administrative cost of collection (probably at least 5 percent), but also distorts other decisions on the allocation of resources and effort.

A third advantage is that congestion levies encourage efficient use of the port facilities. If the congestion charges are raised on the appropriate facilities, this may be associated with a larger throughput of traffic. Even if port throughput is reduced, however, the diversion to other ports or to other uses will be an efficient use of resources.

A fourth advantage of congestion pricing, as distinct from nonprice allocative systems, is that the port authority receives the rent rather than the lucky recipient of the priority or other licensing authorization. Such distributions of wealth among agents, importers, and steamship owners is unlikely to be consistent with the normal canons of equity.

Mobilization Function

At the same time as a price increase rations a scarce commodity, it may also stimulate an increase in supply. Higher-cost producers may now enter the market. The higher price may also induce other firms to increase their investment and scale of production. For example, an increase in energy prices relates to both the rationing function of prices and the mobilization function of prices by inducing more exploration for energy sources and more investment.

Of particular importance in a developing economy is the supply response of agriculture to a change in price. This is considered in Selection 2.7.

3. An interesting analogue is the congestion levies for urban areas. The case for such levies, argued in many studies (for example, A. A. Walters, Economics of Road User Charges, World Bank Staff Occasional Papers, no. 5 [Baltimore, Md.: Johns Hopkins Press, 1968]), remained speculative until the area licensing scheme was introduced in Singapore in 1974. Then, although many mistakes were made, the undoubted success of that scheme showed that the theoretical and empirical studies had predictive relevance. The same may be true of port congestion charges.

4. For the tale of the tub, see Walters, Economics of Road User Charges.
2.7 Price Responsiveness in Agriculture

If and when a positive price policy does become a part of growth policy, it has three functions: (1) to accelerate the growth of agricultural output as a whole; (2) to accelerate or decelerate the growth of the output of individual crops or, in the context of planning, to steer the crop mix according to targets; and (3) to secure adequate increases in the marketed supply of food crops in countries in which a large part of output is retained by the peasants for home consumption.

Adjustments in the crop mix and variations in the marketed supply of food crops are related to the movement of intercrop price relatives, while the growth of aggregate output is related to the movement of the agricultural price index relative to the movement of the prices of manufactured inputs and consumption goods purchased by the peasants. Thus the particular prices to be regulated depend on the relative importance of these objectives.

In order to formulate an effective price policy we need reliable empirical knowledge about the degree of responsiveness of supply (in the three meanings distinguished above) to the relevant relative price movements. The responsiveness will, of course, be different in different milieus—which means that elasticities of supply must be estimated separately for different regions and periods. In the three following sections I have attempted to summarize the currently accessible information about the orders of magnitude of the responsiveness of supply (in the three meanings distinguished above) to the relevant relative price movements. The responsiveness will, of course, be different in different milieus—which means that elasticities of supply must be estimated separately for different regions and periods. In the three following sections I have attempted to summarize the presently accessible information about the orders of magnitude of these elasticities seem to have in some low-income countries.

Responsiveness of Single-Crop Acreage

Considerably more work appears to have been done on the responsiveness of the acreages of individual crops to relative price movements than on market supply or aggregate output functions. The price elasticities of acreage are considered to be good (minimum) approximations of the elasticities of output on the assumption that when the acreage in a crop is varied other inputs can be varied pari passu, and over the relevant ranges of the production functions returns to scale are not diminishing.

Some recent estimates of the elasticities of acreage of individual crops are shown in table 2-3. The pattern of acreage responses which emerges can be summarized in groups. The major crops seem to be classifiable into low-response, medium-response, and high-response groups, defined as crops with short-run acreage elasticities in the ranges 0-0.1, 0.1-0.4, and 0.4-0.7, respectively:

<table>
<thead>
<tr>
<th>Crop</th>
<th>Elasticity Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>0.0-0.1</td>
</tr>
<tr>
<td>Wheat</td>
<td>0.1-0.4</td>
</tr>
<tr>
<td>Maize</td>
<td>0.4-0.7</td>
</tr>
<tr>
<td>Barley</td>
<td>0.0-0.1</td>
</tr>
<tr>
<td>Gram</td>
<td>0.0-0.1</td>
</tr>
<tr>
<td>Rubber</td>
<td>0.0-0.1</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>0.0-0.1</td>
</tr>
</tbody>
</table>

In view of the factors constraining peasant behavior in dual agriculture, this pattern is rationally accountable. The subsistence grain crops—wheat, barley, maize, rice, millets, and gram—respond little to relative price variations. The response of rubber tapings by smallholders is also low (and even lower for estate rubber). Crops grown both for subsistence and for sale—wheat, rice, and maize in surplus areas, and sugarcane—fall in the medium-response range. And the commercial fiber crops have the highest acreage elasticities.

If we compare the responses in subsistence and developed farming in tables 2-3 and 2-4, we find that the price elasticity of wheat and rice acreage in poor countries equals that of grain in the United Kingdom and of maize in the United States; and the elasticity for cotton in India, Pakistan, or Egypt turns out to be about twice that in the United States.

The condensation of available results in this way is subject to many qualifications suggested by the detailed studies from which they have been extracted—in addition to the usual caution regarding the varying degrees of reliability of the data used and the short time periods covered. First, the elasticities are not strictly comparable, for some of them are derived from simple regressions with relative price alone as the determining variable, others from multiple regressions in which the effects of other relevant shifter variables have been netted out. The price coefficients of simple regressions are likely to be underestimated. Second, the relative price variables have been defined in many different ways. The deflating indices sometimes include the prices of a large number of commodities, sometimes the prices of a few close production substitutes, and sometimes the price of only one substitute crop. In general, the estimated responsiveness is likely to be greater, the smaller and more appropriate the set of substitute commodities whose prices are included in the price

deflator. Third, in some cases it is a smaller region where the crop is important in the product mix. Again, the coefficients are likely to depend on whether the crop is the dominant subsistence crop or a relatively commercial crop in the region covered. Fourth, the time period of response is important. When a Nerlovian adjustment mechanism is specified, we are likely to get larger coefficients than when it is not—provided that adjustment lags are real. And, of course, the long-run elasticities will be different from, and larger than, the short-run elasticities.

These considerations explain, in particular, why wheat, maize, and rice appear in our summary in both the low-response and the medium-response categories. The elasticity of wheat acreage in postwar West Pakistan is in the medium range, but in prewar undivided Punjab it was very low. The maize elasticity in prewar undivided Punjab is in the medium range because maize was not a major subsistence crop in the Punjab. But in the Philippines it turns out to be very low. If, however, we look at the elasticities in particular regions of the Philippines, rather than at the elasticity for the country as a whole, they turn out to be in the medium range in many regions. The elasticity of rice in Pakistan, where it is the dominant subsistence crop, is very low; but in undivided Punjab, and in Indonesia, it is in the medium range.

It is remarkable that we do get some credible picture of the patterns of responsiveness of different crops to price movements in different regions. In the available studies there is only one case in which we get a significant negative price elasticity: the case of sorghum in undivided Punjab in the prewar period. The reason seems to be that it was an inferior feed crop. There are also indications of negative coefficients for maize and rice for some regions of the Philippines. In the case of rubber we get the apparently surprising result that, although it is a fully com-

Table 2-3. Estimated Price Elasticities of Acreage of Specified Crops, Less Developed Countries and Regions

<table>
<thead>
<tr>
<th>Crop and country or region</th>
<th>Period</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Short-run</td>
</tr>
<tr>
<td>Rice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punjab (India-Pakistan)</td>
<td>1914-45</td>
<td>0.31</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1948-63</td>
<td>0.05</td>
</tr>
<tr>
<td>Indonesia (Java and Madura)</td>
<td>1951-62</td>
<td>0.30</td>
</tr>
<tr>
<td>Philippines</td>
<td>1947-63</td>
<td>—b</td>
</tr>
<tr>
<td>Ilocos</td>
<td>1954-64</td>
<td>0.22</td>
</tr>
<tr>
<td>C. Luzon</td>
<td>1954-64</td>
<td>0.13</td>
</tr>
<tr>
<td>S. Tagalog</td>
<td>1954-64</td>
<td>0.24</td>
</tr>
<tr>
<td>E. Visayas</td>
<td>1954-64</td>
<td>0.13</td>
</tr>
<tr>
<td>Cagayan</td>
<td>1954-64</td>
<td>—b</td>
</tr>
<tr>
<td>Wheat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punjab (India-Pakistan)</td>
<td>1914-43</td>
<td>0.08</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1944-59</td>
<td>0.20</td>
</tr>
<tr>
<td>Maize</td>
<td></td>
<td></td>
</tr>
<tr>
<td>India-Pakistan</td>
<td>1914-43</td>
<td>0.23</td>
</tr>
<tr>
<td>Philippines</td>
<td>1911-41</td>
<td>0.02</td>
</tr>
<tr>
<td>Philippines</td>
<td>1947-64</td>
<td>0.07</td>
</tr>
<tr>
<td>Ilocos</td>
<td>1953-63</td>
<td>0.07</td>
</tr>
<tr>
<td>C. Luzon</td>
<td>1953-63</td>
<td>—b</td>
</tr>
<tr>
<td>S. Tagalog</td>
<td>1953-63</td>
<td>0.42</td>
</tr>
<tr>
<td>E. Visayas</td>
<td>1947-63</td>
<td>0.40</td>
</tr>
<tr>
<td>W. Visayas</td>
<td>1947-63</td>
<td>0.03</td>
</tr>
<tr>
<td>Cagayan</td>
<td>1953-63</td>
<td>—b</td>
</tr>
<tr>
<td>Bicol</td>
<td>1953-63</td>
<td>0.16</td>
</tr>
<tr>
<td>Mindanao</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. &amp; W.</td>
<td>1953-63</td>
<td>—b</td>
</tr>
<tr>
<td>N. &amp; E.</td>
<td>1947-63</td>
<td>—b</td>
</tr>
<tr>
<td>Barley</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punjab (India-Pakistan)</td>
<td>1914-45</td>
<td>0.39</td>
</tr>
<tr>
<td>Millets</td>
<td></td>
<td>0.09</td>
</tr>
</tbody>
</table>

(Table continues on the following page.)
<table>
<thead>
<tr>
<th>Crop and country or region</th>
<th>Period</th>
<th>Short-run</th>
<th>Long-run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punjab (India-Pakistan)</td>
<td>1914-45</td>
<td>-0.33c</td>
<td></td>
</tr>
<tr>
<td>Sorghum</td>
<td>1914-43</td>
<td>-0.58</td>
<td></td>
</tr>
<tr>
<td>Sugarcane</td>
<td>1915-43</td>
<td>0.34</td>
<td>0.60</td>
</tr>
<tr>
<td>Cotton</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India-Pakistan</td>
<td>1922-43</td>
<td>0.59</td>
<td>1.08</td>
</tr>
<tr>
<td>India</td>
<td>1948-61</td>
<td>0.64</td>
<td>1.33</td>
</tr>
<tr>
<td>Pakistan (8 districts)</td>
<td>1933-58</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>1935-62</td>
<td>0.50d</td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td>1900-38</td>
<td>0.4-0.66</td>
<td></td>
</tr>
<tr>
<td>Jute</td>
<td>1911-38</td>
<td>0.46</td>
<td>0.73</td>
</tr>
<tr>
<td>India-Pakistan</td>
<td>1911-38</td>
<td>0.68</td>
<td>1.03</td>
</tr>
<tr>
<td>Bengal (India-Pakistan)</td>
<td>1893-1938</td>
<td>0.57-0.65</td>
<td></td>
</tr>
<tr>
<td>India-Pakistan (Bengal, Bihar, Orissa)</td>
<td>1893-1938</td>
<td>0.57-0.65</td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>1931-53</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>1948-63</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Rubber</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estates</td>
<td>1953-60</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Smallholders</td>
<td>1953-60</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Estates</td>
<td>1951-61</td>
<td>-0.02i</td>
<td></td>
</tr>
<tr>
<td>Smallholders</td>
<td>1948-61</td>
<td>0.12f</td>
<td></td>
</tr>
<tr>
<td>Estates</td>
<td>1954-61</td>
<td>0.03c,h</td>
<td></td>
</tr>
<tr>
<td>Smallholders</td>
<td>1953-60</td>
<td>0.34b</td>
<td></td>
</tr>
</tbody>
</table>

a. For summer and winter crops combined; elasticity for summer crop alone, 0.12.
b. Negative coefficient indicated.
c. Not significant at 10 percent level.
d. Based on simple calculations from yearly variations.
e. Based on year-to-year arc elasticities.
f. Based on annual data.
g. Based on monthly data.

### Responsiveness of Market Supply

Besides the elasticities of acreage (output) of individual crops we need to know, if possible, the elasticities of their marketed supply. In the case of crops which are wholly or almost wholly marketed, the elasticities of output and market supply can be regarded as approximately equal. But in the case of crops, a substantial part of whose output is retained by the peasants for home consumption, the responsiveness of the marketed supply must be measured separately from the responsiveness of output.

It is clear that where the price elasticity of output itself is negative, the elasticity of market supply will also be negative. But this is an extreme case. As we saw in the previous section, the empirical studies available to date have not thrown up many cases of...
Table 2-4. Estimated Price Elasticities of Acreage of Specified Crops, United Kingdom and United States

<table>
<thead>
<tr>
<th>Country and commodity</th>
<th>Period</th>
<th>Elasticity</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Short-run</td>
<td>Long-run</td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grain (all grains)</td>
<td>1924-39</td>
<td>0.12</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>1924-39</td>
<td>0.33</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Barley</td>
<td>1924-39</td>
<td>0.63</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>Oats</td>
<td>1924-39</td>
<td>0.11</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1946-58</td>
<td>0.18</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td>1907-40</td>
<td>0.39</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Peas</td>
<td>1938-58</td>
<td>0.65</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>1909-32</td>
<td>0.48</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>1909-32</td>
<td>0.10</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Cotton</td>
<td>1909-32</td>
<td>0.27</td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>

negative elasticities of output, although it is possible that extreme cases of perversity exist precisely in the zones where research workers do not reach and cannot get the data for estimating elasticities.

The interesting cases are likely to be those in which the output response is positive and the market supply response negative. This situation can exist if an increase in the relative price of a subsistence crop causes a greater increase in consumption than in output. We cannot at present establish empirically the frequency with which such situations may exist, for in order to do so we would need reliable time series data about acreage, output, prices, and market sales for specified regions. Market sales data for food crops are seldom available, even where other data are. Therefore market supply functions cannot be estimated directly. An attempt has been made to devise an indirect method of estimating the approximate range within which the elasticity of the market supply of a subsistence crop may be expected to lie if the parameters that determine it lie within certain estimated ranges. This attempt showed only that the range as well as the magnitude of the elasticity of market supply depends critically on the combination of the magnitudes of these other parameters. With one set of their values, the elasticity of market supply can be negative or positive but very low; with another set, it can be positive and high.

No one can assert that the elasticity of market supply must be negative or positive in most cases. Both logically and empirically it can be positive in some and negative in other regions; and even in the same region it can be positive for aggregate market supply and negative for particular subclasses of farmers. Also, the elasticity must, of course, differ in the case of different crops.

All the important logical possibilities cannot be discussed here. But it may be useful to set out the quantitative framework that determines the outcome. Suppose the output Q of a crop is divided between consumption C and sales M. Then differentiating the identity M = Q - C with respect to relative price P we can write the elasticity of market supply

\[ e = rb - (r - 1)d \]

where \( r \) is the reciprocal of the sale ratio \( M/Q \); \( b \) is the price elasticity of output; and \( d \) is the total price elasticity of home consumption. The reciprocal of the sale ratio is positive and greater than unity because the sale ratio is positive and less than unity. Therefore \( r \) and \((r - 1)\) will be positive. Considering the influence of the different variables on the sign of \( e \), we can distinguish many possibilities.

1. If \( b \), the elasticity of output, is negative and \( d \), the elasticity of consumption, is positive, \( e \) will be negative.
2. If \( b \) is negative and \( d \) is also negative, \( e \) will be positive so long as \(|rb| < |(r - 1)d|\).
3. If \( b \) is positive and \( d \) is negative, \( e \) will be positive.
4. If \( b \) and \( d \) are both positive, \( e \) will be positive so long as \( rb > (r - 1)d\).

Thus in two cases the sign of \( e \) depends only on the signs of \( b \) and \( d \). But in the other two cases the sign of \( e \) depends on the magnitudes as well as the signs of \( r \), \( b \), and \( d \). Assuming plausible ranges of \( r \), \( b \), and \( d \), we can measure the approximate range of \( e \) for a given crop in a given area. The value of \( b \) can be taken from output (acreage) functions, and \( r \) from family budget data. But it is difficult to compute \( d \). For what is needed in equation (1) is the total price elasticity of home consumption—including the effect of a price change on the income of the farmer. In a multicrop economy, this effect depends on the
proportion that the subsistence crop under study contributes to the total income of the peasant and on how the change in income following a change in the price of the crop is defined.

The definition of the income of a peasant unit or a peasant community is itself a difficult task. Income is usually derived from many sources: crop production, cottage industry production, off-farm labor, and the like. Some of it is cash income, but much of it is income in kind. The proportion of owned and hired resources, and hence the paid-out cost, varies from unit to unit. The existing institutional arrangements make it very difficult to price resources. There are serious objections to the valuation of retained output and sold output at the same price, and to the valuation of all labor at the wage rate which can be earned if and when off-farm work is available. In any case, measures of peasant income defined in alternative ways are not accessible, and we do not know the measure of income with which the consumption of peasants is most closely correlated. In short, we do not have any reliable price-income and income-consumption relationships at the present time. Until these relationships can be clarified and estimated we cannot have even indirect estimates of the price elasticity of the marketed supply of subsistence crops.

We can, however, think of another method of getting at the price elasticity of market supply, which avoids some of these difficulties. The price elasticity of output can be estimated as usual. We can also estimate from cross-section data the market supply-output relation, and compute the output elasticity of market supply. The price elasticity of market supply can then be calculated as the product of the price elasticity of output and the output elasticity of market supply, assuming that a cross-section estimate of the latter approximates a time series estimate at least for years close to the cross-section year:

\[ e = bE_{NCQ} \]

Two or three attempts have been made recently to estimate the relationship of market supply to output from cross-section data for some parts of India, Pakistan, and the Philippines. The output elasticity of market supply of wheat was found to lie between 1.04 and 1.6 in many villages in India. If we take the (marginal) price elasticity of output to be as low as 0.1 or 0.2, the price elasticity of market supply should be expected to lie between 0.104 and 0.32. Khan and Chowdhury found the output elasticity of the market supply of food crops (grains, pulses, and vegetables) in West Pakistan to be approximately 1.6. Assuming again the price elasticity of output to be 0.1 or 0.2, the price elasticity of market supply should be between 0.16 and 0.32.

In India the relation between the market supply and the output of wheat turned out to be linear in some cases and nonlinear in others—with a positive slope and a negative intercept. Since the output elasticity of market supply is usually found to be greater than unity, we can say that the price elasticity of market supply should normally be greater than the price elasticity of output. This has the important implication that even if we cannot estimate the market supply function directly, we can assume that wherever market supply is positive its elasticity is not likely to be less than the elasticity of output, which we can estimate. In other words, if output increases in response to any price increase, sales may be expected to increase more than proportionately with output; but if output falls for any reason, market supply will fall more than proportionately. Here we see why the supply of food in urban markets fluctuates more than the output of food.

With regard to the marketed surplus of agricultural commodities as a group, we have no time series studies at present; and we have only one important cross-section study, for India, which revealed a tendency for the surplus-output ratio to fall and then to rise as size of holding increases. The explanation offered for this phenomenon has been that the very small farmers have to engage in distress sales to meet their onerous payment obligations; farmers with a little more land reduce their sale ratio and improve their consumption; but the sale ratio of farmers having a holding of more than a certain critical size increases normally with the holding size. Cross-section studies for other peasant societies are urgently needed to establish the degree of generality of this finding.

**Responsiveness of Aggregate Output**

So far the least amount of work has been done on the response of aggregate output (of all agricultural products or all crops) to the relevant relative price index, namely, the terms of trade of agricultural production. This gap in empirical research is especially regrettable because knowledge of this response is crucial for the formulation of a price policy for agricultural development. A priori expectations that were popular some time ago about the existence of backward-sloping supply curves in peasant economies have not been confirmed by recent empirical work in respect of the response of the acreages (output) of individual crops in many parts of the peasant world. But it is possible to argue that the responsiveness of individual crops to relative price movements may be positive, and still the responsiveness of aggregate output may be zero or negative. For peasants may switch land or other resources available to them between different crops (products) at the margin; but they cannot, with the same ease, add to the sum total of resources committed to agriculture or shift them out of agriculture, or increase their productivity when
prices move up and down. In other words, peasants may show some allocative rationality and yet be unable to increase (or decrease) their total output substantially in response to a change in their terms of trade.

Now, there is much obvious truth in this view. But when we study empirical reality the truth is, as usual, very complicated. First, consider the transport context of the agriculture of different regions. For areas that are almost isolated for want of transport and communications facilities (with all that this implies) the very question of responsiveness is meaningless. To say that the agriculture of an isolated area is not price responsive is to say something trivially obvious. The question of responsiveness is meaningful only for open regions where a critical minimum of transport development has occurred and the concomitant processes of monetization and commercialization have been set in motion. In the case of such regions the price responsiveness of total output or the lack of it has to be established empirically. There is no point in asserting that in all these regions output will or will not be price responsive. And the degree of responsiveness should be expected to differ from region to region.

Second, there are many parts of the world where a single crop or two crops dominate the product mix. Here the distinction, and the possibility of asymmetry, between the responsiveness of output of these crops and total output is unimportant. If the output of the dominant crops is positively responsive at all to increases in their prices, and the output of minor crops does not fall, we can infer that total output is also positively price responsive at the margin. The same inference can be drawn if we find that the output of minor (mainly commercial) crops is positively price responsive to increases in their prices while the output of dominant (mainly subsistence) crops is maintained.

Third, in many sparsely populated parts of the world where occupiable land has been available, more and more of it has been brought under cultivation to meet not only the food requirements of a growing population but also the pressure of demand for commercial crops (usually for export). Thus aggregate output has grown in response to price increases reflecting both types of pressure.

In densely settled multicropping agriculture, of course, resource inelasticity (particularly the inelasticity of the supply of land) is likely to be the greatest. Therefore the same total area of land may be reallocated by the peasants between different crops in response to relative price movements, but total output may not vary much because of the limitation of land. In other words, the responsiveness of aggregate output in dense areas depends on the potential for increasing per-acre productivity realizable by the peasants themselves; but with their given knowledge and material resources they cannot realize much of it. New knowledge and additional resources have to be brought in by government or other outside agencies before the potential productivity increases can materialize.

Also, some increases in yield may materialize because of more intensive application of labor to irrigation and land improvement, more careful husbandry, and greater use of farm compost, if prices are favorable. Thus we may find some positive yield responses to favorable price movements at least in the case of commercial crops, even in a milieu of very limited knowledge and resources.

Fourth, when governments or outside agencies do begin to increase the peasants’ stock of knowledge and material resources, price responsiveness may reflect itself in a new form. The development agencies may find that the rate of absorption of these inputs depends inter alia on their return-cost ratios, which in turn depend on product-to-input price ratios. The evidence clearly suggests that even when a government is doing its best to restructure agriculture institutionally, and to expand the availability of knowledge and inputs, it finds that output remains stagnant until these measures are supplemented by a positive agricultural price policy. The engines of extension and institutional reform do not raise output fast enough without the steam of price incentives. Such evidence is also an important indirect indication of the positive price responsiveness of total output at the margin.

In measuring the contribution of price movements to agricultural growth we must not lose sight of the fundamental truth that the transformation of traditional agriculture is primarily a techno-organizational episode. The transformation cannot be brought about only or mainly by price movements. However, the techno-organizational effort can be retarded or accelerated by price movements. Favorable price movements can speed up the diffusion of innovations, the absorption of new inputs, the utilization of idle capacity, and even institutional adjustments. Unfavorable movements can slow down or arrest all these processes.

If a developing country suffers from a shortage of savings, it is important to recognize how the level of interest rates may affect the level of savings. Selection 2.8 shows the negative effect on the volume of loanable funds of a policy of low interest rates. The reforms in the Korean economy provide another instructive example of how prices may stimulate the mobilization of resources to finance development (Selection 2.9).
2.8 Interest Rates on Agricultural Loans

Interest Rate Policies in Developing Countries

A persistent pattern of low interest rates is found in credit institutions. Table 2-5 provides a relatively comprehensive survey of such interest rates. There are several respects in which these rates can be regarded as being too low.

1. The rates charged are usually comparable to or lower than the rates charged in the more advanced, capital-rich countries. Most of those in table 2-5 are within the range of 6 to 12 percent a year, and there are other reported instances of still lower rates such as 3 or 4 percent. (The few cases in the table of rates higher than 12 percent a year are almost all in countries with high rates of inflation.) Most of these rates would appear to be low. Such rates of interest do not come near to the actual market rates, or to most of the shadow prices for capital calculated in planning exercises in these countries.

2. In the majority of cases, the rates of interest charged on public agricultural credit are lower by approximately 3 percent than the prevailing commercial bank rates in the same countries. Where there are small-farmer-specific credit programs, most of the interest rates charged are lower than those applied to other agricultural credit. When low rates of interest imply a subsidy, preferential rates imply an additional subsidy for which a specific justification should be provided.

3. Given the existing national rates of inflation, the real rates of interest charged by most of these programs are even lower than the nominal rates quoted and are sometimes negative—see table 2-5. (A negative real rate of interest results from a rate of price increase that is higher than a nominal rate of interest.) In several Latin American countries, real rates have been substantially below zero for long periods of time. Real rates of interest are more relevant than nominal values in terms of their impact on the real value of lenders' portfolios, on the actual size of the subsidy captured by borrowers, and on economic behavior in general.

4. The rates of interest charged by small-farmer credit institutions are low in that they tend to generate a demand for credit greater than the supply of institutional funds. When demand exceeds supply, “rationing” of credit becomes necessary. The availability of credit for small farmers is then influenced by the rationing mechanism chosen to allocate loans among farmers.

5. Another sense in which the rates charged are low is that they often do not cover the average cost of delivering credit to small farmers. The financial difficulties that a lending institution would have to face if it expanded its services to small farmers under these circumstances will, in turn, influence adversely the choice of the rationing mechanism that the under-equilibrium rate requires.

6. Financial institutions must pay even lower rates of interest on deposits and financial instruments for the collection of savings than those charged for loans. The low level of such rates has inhibited the organization of savings programs and thus curtailed a potential source of loan capital for these lending institutions.

Reasons behind These Policies

There are two types of reasons why low interest rate policies have been adopted for the formal financial markets. One type reflects the general nature of monetary and financial policies in developing countries. The other type of reason is related to problems that are thought to arise in connection with agriculture in general, and with small farmers in particular. MONETARY AND FINANCIAL POLICIES

At the macroeconomic level, most developing countries have not pursued a flexible interest rate policy. Interest rates have been excluded as an active policy instrument from prevalent theories of development. The level of interest rates has been legally fixed in most cases, and the number of instances in which these fixed rates have been revised is very small. This is particularly striking in countries that have consistently experienced high and widely varying rates of inflation. Policymakers have been more preoccupied with the nominal level of the rates and with the moral implications of “usury” than with the possibility of enlisting the rate of interest as a powerful instrument for development.

Development theorists, indeed, have frequently argued that strategic economic activities are not responsive to the signals that originate in the level and structure of interest rates. Their proof usually comes from some econometric study of the impact of interest rates in developed countries, and the theoretical underpinning usually derives from short-run Keynesian theories of unemployment and depression in advanced countries that are inapplicable in developing countries. Rigorous analysis, however, shows that interest rates are crucial determinants of the processes of capital accumulation and of allocation of resources.
Table 2-5. Interest Rates Charged by Agricultural Credit Institutions

(annual percentage rate)

<table>
<thead>
<tr>
<th>Location</th>
<th>Nominal rates $^a$</th>
<th>Real rates $^b$</th>
<th>Rate of inflation (1967–70) $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Ghana</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Ivory Coast</td>
<td>10 $^b$</td>
<td>6 $^b$</td>
<td>4</td>
</tr>
<tr>
<td>Kenya</td>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Morocco</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Nigeria</td>
<td>6 $^b$</td>
<td>-2</td>
<td>8</td>
</tr>
<tr>
<td>Sudan</td>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Tunisia</td>
<td>6 $^b$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Uganda</td>
<td>12</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afghanistan</td>
<td>9</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>12</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>India</td>
<td>9</td>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>Indonesia</td>
<td>14</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Iran</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Jordan</td>
<td>7</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Malaysia</td>
<td>18</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>Pakistan</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Philippines</td>
<td>12</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>So. Korea</td>
<td>16</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>So. Vietnam</td>
<td>30</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>12</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Taiwan</td>
<td>10</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Thailand</td>
<td>11</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Latin America</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bolivia</td>
<td>9</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Brazil</td>
<td>15</td>
<td>-7</td>
<td>22</td>
</tr>
<tr>
<td>Chile</td>
<td>14</td>
<td>-16</td>
<td>30</td>
</tr>
<tr>
<td>Colombia</td>
<td>12</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Ecuador</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>El Salvador</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Honduras</td>
<td>9</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Mexico</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Peru</td>
<td>10</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

a. Nominal rates are averages of the reported rates charged on various types of agricultural loans by one or more institutions in given countries. Generally the rates in private banks are higher than those in public institutions by about 3 percentage points.

b. Real rates were obtained by subtracting from nominal rates the average annual rate of increase in the consumer price index over 1967–70, the period for which most nominal rates were reported.

c. Not an annual rate; a commission of 10 percent is charged on loans of varying durations less than a year.

Source: World Bank data.

and therefore lie at the very heart of the development problem. Their influence is particularly crucial during periods of rapid changes in technology, with new production frontiers and consumption opportunities. In the few cases in which more reliable empirical information is available for developing countries, significant changes in consumption, savings, investment, the demand for liquid assets, production, and choice of techniques have been induced by substantial changes in the real level of the rate of interest.

We must consider that interest rates have had an important and widespread influence on economic behavior even though this influence has so far been examined carefully in only a few developing countries, and behavioral functions cannot yet be quantified. By fixing artificially low rates of interest, pol-
icymakers have attempted to render them inoperative; resource allocation could then be influenced by allocation procedures chosen by the policymakers. But while interest rates have been inhibited from serving as efficient economic signals, they have nevertheless influenced decisions and affected the evolution of economic life. This influence has seldom been conducive to either efficiency or equity.

Interest rates are relative prices. In the most fundamental sense, they are the price of the future in terms of the present. The level at which the rates of interest are set influences consumption, savings, and investment behavior. Low rates of interest tell savers not to bother with saving, that the future is amply provided for, that now is the time to consume. High rates of interest tell consumers that sacrifices in present consumption will be highly rewarded. Low rates of interest on loans tell investors that investment goods are plentiful for use in the production of consumables. High rates of interest tell investors that capital goods are scarce and have to be economized.

Artificially low rates of interest motivate consumers to bid resources into consumption markets, but these low rates also encourage investors to bid resources away from consumption. This conflict is resolved by administrative decisions on who will get loans and who will be denied them. Interest rates are, in this way, deprived of their role as a device that requires potential spenders to pit their spending options against others in a way that reflects productivity and consumer time preferences. Rates of interest are also the prices relevant in financial markets. As a deposit rate paid on financial assets, interest rates affect the allocation of saving among various types of assets, some of them more socially productive than others. Savings incorporated in gold, jewelry, luxurious housing, and any other form of idle inventories have little social productivity. Low rates of interest not only direct savings towards those assets, but also into inventories of commodities which may involve costs of depreciation and handling. Rates of interest that underprice savings not only put them to inferior uses—other than productive investments—but also divert them to financial markets abroad. In a fragmented capital market, the release of resources from inferior uses is as important as the accumulation of capital per se. A policy of higher interest rates could improve both.

As the price of a factor of production, loan rates of interest affect the relative factor intensity of the productive processes undertaken. Low rates of interest charged on loans direct resources into more capital-intensive methods of production which can increase unemployment in capital-poor, labor-abundant countries. This could take the form of premature mechanization of agriculture in the presence of heavy rural unemployment, for example. In sum, rates of interest affect the demands for and supplies of goods, assets, and factors of production over time, and in this way influence the rate of growth of the economy, price stability, and the distribution of income. Inappropriate interest rate policies distort capital markets and reduce their contribution to economic development. The low rates of interest charged in small-farmer programs are a reflection of these more general policies.

CREDIT AND AGRICULTURAL SECTOR PROBLEMS

The rates charged by small-farmer credit institutions are usually preferential. Several arguments have been advanced to justify this policy.

1. In many cases, preferential rates have been justified on the grounds that small farmers have been exploited by informal lenders and by market intermediaries in general. A concern for the welfare of poor farmers is probably the most influential cause for adoption of preferential low interest rates by institutions intended to serve them. But most of this low-interest credit does not actually get to the small farmer. Subsidized credit which is captured by larger farmers only creates conditions disadvantageous to small farmers to the extent that its beneficiaries will use the funds borrowed from the lending agency to buy additional land, or to displace small farmers in the purchase of scarce inputs—that is, it adds to the large farmers’ access to resources.

2. Prevalent interest-rate policies have been based, in other instances, on the claim that traditional farmers need special inducements to use highly productive inputs, and that only highly subsidized credit will induce them to adopt modern technologies. A more sophisticated version of this justification for subsidy is based on differences between the private costs and benefits and the social costs and benefits of an activity such as a project to promote new farming techniques. Because of this divergence, the activity does not expand as much as would be socially desirable in the absence of a subsidy. Subsidized rates of interest have been justified on these grounds. Further analysis is necessary, however, to determine whether in fact such a divergence between private and social costs and benefits exists, and what is its nature.

   a. In some cases, lack of private profitability simply reflects low social returns to an innovation. Reluctance on the part of small farmers to borrow funds and adopt the new technology thus reflects a lack of profitability that should not be countered with a subsidized rate of interest. In these circumstances, a subsidized rate of interest will lead only to inefficiency.

   b. In other instances, the lack of profitability of the innovation highlights unavailability of certain inputs, the existence of a bottleneck, or the absence of a market for the product. Such obstacles will not
often be removed by an expansion in the volume of credit. A subsidized rate of interest will not create a missing market; a high enough subsidy, however, can transform an unprofitable activity into a profitable one, despite the continued weakness of the market. Subsidized credit, then, only perpetuates the imperfection. By concentrating attention on the credit link of the process, measures dealing directly with the imperfection are not taken.

c. Reluctance to borrow, even at relatively low rates of interest, might reflect the absence of an appropriate technology. The distribution of credit at subsidized rates of interest might induce farmers to accept the "gift" and to use the funds for consumption, but it will not make the missing technology available. Rates of default will then be high. The attempt to induce farmers to use credit by means of a subsidized rate of interest diverts attention from the basic weakness of the available technology and from the measures needed to improve it.

d. There may nevertheless be circumstances in which, because of the learning processes involved and the demonstration effects generated, relatively sophisticated "infant industry" arguments could be advanced. But even in the cases for which some sort of subsidy could be justified, it is not obvious that the most efficient policy instrument is an interest-rate subsidy. As a general rule, subsidies should be given directly at the source of an externality. In most instances relevant to agricultural development, the learning processes and externalities involved are not a direct function of the use of credit. A subsidized rate of interest might expand the activity upon which the learning process depends, but it will also cause additional distortions. For instance, it could artificially increase the capital intensity of production.

3. Low interest rates have also been justified as a mechanism for income transfers to small farmers, and to offset fiscal, foreign exchange, and pricing policies that have adversely affected the agricultural sector. Low rates of interest are of course advantageous to the individual farmers who are able to borrow at such rates. Unfortunately, the mechanisms of credit allocation induced by low rates of interest have consistently discriminated against small farmers, as will be demonstrated below, and have restricted their access to credit. Little consolation can be gained, therefore, from the contemplation of the very small proportion of small farmers who have received credit at low subsidized rates, while the large masses of them continue to depend on a fragmented informal market in which high rates of interest are charged.

Thus, the use of credit as a mechanism for income transfers to poor farmers is very inefficient. Furthermore, the perception by the farmer that he is dealing with a "welfare" program designed to transfer resources to poor farmers will induce a reluctance to repay his loans, even if he could, particularly if he realizes that more powerful members of the community are not paying back either. Lack of repayment, of course, effects an additional income transfer but in an erratic fashion not controlled by the lender. If defaulting is a function of political power and other status variables, much of the income transfers channeled through the credit program will be delivered to nonmembers of the target population, while the program is kept going by the many small farmers who do repay. In any case, this kind of unacknowledged rural welfare burden is unfair to the administering lending agency. The institution is forced into a dependence on outside funds, while its performance will still be judged as a failure by the normal criteria applied to credit institutions.

In summary, the policy of low and preferential interest rates is generally justified as a means of helping small farmers, but analysis of the implications of such a policy shows that these efforts are misdirected. The policies adopted have actually misfired, and have in most cases led to a more limited access to credit for small farmers than would otherwise be the case. We shall now examine why this result is probable—indeed inevitable.

**Negative Impact of Low-Interest Rate Policies**

As a mechanism for improving the access of small farmers to credit, a policy of low interest rates presents important disadvantages because it actually reduces the total volume of credit flowing into the small-farmer sector. A higher rate of interest charged on loans would increase the amount of resources that an institution has available for loans in the following year. A low rate of interest depletes the credit agency's resources and reduces the volume of loanable funds that it manages. It also closes off the potential loan funds that it might attract if it could pay higher interest on savings deposits.

Low rates of interest charged by formal lenders may also reduce the funds available to small farmers from the informal market when the formal institutions offer too much competition for informal lenders. If low interest attracts the majority of large, low-cost and low-risk customers to a credit agency, the informal lenders are left to serve the most costly and risky potential borrowers, which may discourage their operation. If the interest ceilings imposed on formal lenders were to be enforced on the informal lenders, as is sometimes attempted, this would further reduce the supply of funds for small farmers.

**SMALL FARMERS' SHARE OF CREDIT**

Low rates of interest reduce the propensity of formal lenders to service small farmers. All lenders, sooner
or later, no matter what their objectives, are forced to find some kind of balance between revenues and costs. Interest payments constitute the main source of revenues for a credit agency. Low rates of interest generate low revenues, and often these are not sufficient to cover all the costs of delivering credit to small farmers. Given strong pressures on an institution to keep its costs in line with revenues, and given the shape of the cost functions involved, the agency is forced to reduce the proportion of its portfolio devoted to small farmers. By increasing the average size of a loan, the institution can bring down average costs. Where the potential for government subsidies is limited or uncertain, such a cost reduction may be a necessity for survival.

If the rates of interest they are allowed to charge do not cover their average costs, it is not surprising to find that profit-maximizing institutions such as commercial banks are reluctant to serve the rural areas. Losses are also a hazard for public institutions, even though they may attempt to rationalize them in terms of "social" objectives. Public lending institutions cannot be sure that they will permanently receive ample funds from their governments, or from international donors. In any case, the continuous receipt of such outside funds subjects the institution to an undesirable dependency on the political-administrative process of allocating public funds, and leaves it highly vulnerable to political pressures. It is not realistic to assume that small farmers will always have enough political power to maintain a substantial allocation of funds in their favor.

Lack of sufficient and timely appropriations can severely impair the efficient functioning of any subsidy-dependent institution that serves small farmers. The continuity of the program itself, the possibility of achieving its goals, not to speak of the possibility of expanding the program to reach more of the target population, are all threatened by continuing losses. The prospect of a disappearance of its capital due to high and persistent operational losses, an outcome related by several Spring Review papers, leads either to the elimination of the program or to its reorganization away from the service of the small farmer. The danger of erosion of the volume of loanable funds is particularly acute in inflationary environments, since not only is the nominal value of the funds reduced by losses, but the real value of the remaining funds is quickly devoured by rising prices.

All these conditions lead to one result: the lower the interest rate charged on loans, the lower the proportion of the lender’s portfolio, ceteris paribus, that will be devoted to small-farmer credit.

CREDIT RATIONING

Another influence working against small farmers is that the low rates of interest charged tend to generate an excess demand for institutional funds. At the rate of interest set there are more potential clients willing to borrow than those who could be satisfied with the institutional resources available, so that some non-price mechanism for rationing becomes necessary. The lending institution must select the beneficiaries of the credit program, rejecting other potential users of its funds. The lower the rate of interest, the more extensive is the excess demand generated, and the greater will be the reliance of lenders on rationing devices for allocation of their limited funds. The lower the rate of interest and the higher the implied subsidy, the greater will be the motivation of borrowers to exert social and political power in order to capture it. The lower the rate of interest and the greater the potential losses for the credit institution, the more likely it will be to succumb to such political pressures.

The high scarcity value of the limited resources that are being distributed at low official prices creates the conditions for a “black market price.” This price may involve extra payments or bribes, which richer farmers are better able to pay, or it may take the form of political and social pressures on the institution and its officers by influential persons. Such influences on the allocation of credit have led everywhere to a distribution of the credit pie in favor of the larger farmer. There have been instances when the allocation of subsidized funds has been turned into a purely political struggle, as in Bolivia and Morocco.

Another consequence of low rates of interest is that lending agencies overemphasize their reliance on collateral requirements as criteria for borrower selection. This emphasis on collateral results partly because lenders have an increased risk aversion when revenues do not cover costs, and partly because it represents a convenient rationing device for loan allocation in the face of excess demand. Collateral requirements restrict access to credit to those who own sufficient assets to be acceptable to the credit agency, i.e., the wealthier members of the community. In a number of countries many small farmers do not have adequate titles to the land they operate, whether because they are tenants or because title registration is incomplete, especially in the transition from tribal ownership in Africa. It is the larger farmers who have the clearer land titles along with other assets.

Other kinds of rationing devices also discriminate against small farmers. This is the case with the long delays and elaborate formalities in loan approval, which serve to screen out potential borrowers. Timing is crucial in agriculture: if the resources are not available when needed for a particular activity, the activity cannot be undertaken. When credit is being granted too slowly, only those farmers with enough
resources of their own can afford to go ahead with the activity and receive the borrowed funds later.

2.9 Financial Deepening

Financial liberalization tends to raise ratios of private domestic savings to income. Some of the gain may be illusory: savings may be drawn to uses that are counted in national income accounts and away from capital flight in various forms that defy measurement. Some of the gains must result from higher rewards to saving—higher real rates of interest and opportunities for diversifying savers' portfolios of domestic assets. Real growth of financial institutions provides more investors with access to borrowing and gives them an incentive to save and to accumulate the equity that makes borrowing cheaper. Finally, there seems to be a shift of savers' planning horizons to the more distant future and an improvement in income expectations that reduces relatively the attraction of consumption now.

Savings tend to rise in the government sector. When finance is shallow, in decentralized economies, government savings are characteristically low. Government depends on the inflation tax for command over resources, and the typical inelasticity of tax revenues and profits of government enterprise to inflation reduces government savings in the conventional sense. Savings from the foreign sector also respond to liberalization. Capital flight of domestic funds is reversed, and there is easier access to foreign capital markets when distortions of such relative prices as interest rates and foreign exchange rates are corrected. In some circumstances the inflow of foreign funds reaches proportions that are not easy to absorb at a stable domestic price level.

Liberalization permits the financial process of mobilizing and allocating savings to displace in some degree the fiscal process, inflation, and foreign aid. Fiscal technique is backward in lagging economies. Its capacity to draw revenues is strained by demands for government consumption and, in the usual case, for some minimal social security. Additional demands upon it for the purpose of financing investment tend to have effects upon economic efficiency and social equity that offset benefits of capital accumulation. Foreign aid has been, in some important degree, a substitute for domestic savings: the aid gap is partly an expression of excess demand for savings in economies that employ relative price to repress savings of their own. Technocratic planning models may define ranges of per capita income in which developing economies generate shortages of savings and foreign exchange, but correction of distorted relative prices has a remarkable way of making the models irrelevant.

Liberalization opens the way to superior allocations of savings by widening and diversifying the financial markets on which investment opportunities compete for the savings flow. In the repressed economy, savings flow mainly to the saver's own investments: self-finance prevails. In the liberalized economy, savers are offered a wider menu of portfolio choice. The market for their savings is extended; a broader range of selection in terms of scale, maturity, and risk becomes available; and information for comparison of alternatives can be obtained more cheaply. Local capital markets can be integrated into a common market, and new opportunities for pooling savings and specializing in investment are created.

The extended capital markets, where prices in the form of interest rates can be used to discriminate between investment alternatives, seem to be a congenial context for appearance of new investing firms and innovative investment projects. They generate competing proposals for the disposition of savings. In this respect, the contrast with financially repressed regimes can be striking. There, savings flow through narrow channels, which are not subject to the discipline of relative price, into the repetitive projects of established firms, especially government enterprise and traditional trading firms. Too often the search for investment uses of savings is casual and opportunistic, with little reference to comparative rates of return. Financial depth seems to be an important prerequisite for competitive and innovative disposition of savings flows. It is no accident that marginal ratios of investment to output are high where finance is shallow.

Unemployment in lagging economies is partly the result of financial repression. Scarce savings supply labor inadequately with the tools of its trades. To make matters worse, the low interest rates that inhibit savings combine with relatively high minimum-wage rates to guarantee that labor is supplied with the wrong tools. Investment flows to capital-intensive production even though capital is scarce and labor plentiful. The status of labor is not improved by the overvalued exchange rates that conceal the comparative advantage of labor-intensive agriculture and indigenous manufacturing.

Financial liberalization and allied policies tend to equalize the distribution of income. It appears that elasticities of substitution of labor against capital may be high in some lagging economies and that a rise in interest rates and foreign exchange rates relative to wage rates may both raise employment and in-

crease the wage share of income. Liberalization also reduces the monopoly rents that flow from import and other licenses to the few importers, bank borrowers, or, for example, consumers of electric power. Furthermore, the twists in terms of trade against farmers, for example, or against workers, that are applied in a regime of repression to extract profits and savings for investors can be replaced by measures of finance and taxation that achieve as much growth of capital with less abuse of equity and less hazard for political and social stability.

Liberalization and deepening of finance contribute to the stability of growth in output and employment. There can be less stop-go. One reason for this is that some of the more ample savings flow may be used to finance larger international reserves. More flexible foreign exchange rates can also absorb some of the shocks of international trading. If both the current balance on international account and the domestic savings flow respond to treatment, the economy becomes less vulnerable to the ebb and flow of supplier credits and foreign aid. From the standpoint of stability, it is perhaps most important that liberalization can reduce dependence on bursts of inflation and the inflation tax to balance fiscal budgets and can bring monetary variables under discipline.

The name of the policy game in repressed economies is interventionism. Because monetary variables are out of hand, there seems to be a need for price control in detail. Because the exchange rate is overvalued, complex tariff schedules, import licenses, and differentiated export bonuses are put into force. Because savings are scarce, credit is rationed loan by loan. An economy that immobilizes critical relative prices must fall back upon contrivances of interventionism to clear markets. A burden is put upon the civil service that it cannot carry, and the costs in both inefficiency and corruption are high. It is a principal purpose of liberalization to substitute markets for bureaus.

The Origins and Context of Financial Repression

Financial repression has its typical complements in development strategy. It is part of a package. As previously suggested, overvaluation of the domestic currency in terms of foreign monies is another part. Whereas financial policy including inflation reduces real rates of interest and makes savings appear cheap, so cheap that they must be rigorously rationed, exchange rate policy holds down the domestic price of foreign exchange. There is excess demand for both savings and foreign exchange. In the rationing of foreign exchange, preference commonly goes to capital equipment or, once the equipment has been installed, to its spare parts and replacements and to materials that it processes.

Another part of the strategy reduces relative prices for domestically produced primary products. Industrial raw materials are made cheap for the urban industrial establishment, which benefits, too, from low rates of interest and from cheap imports. Foods are made cheap for the urban working class in order to temper its wage demands. Primary products are diverted from export markets, and imports of substitutes have a claim of high priority for scarce foreign exchange when domestic supplies run short.

The strategy produces a dual labor market. A relatively small fraction of the labor force is recruited for capital-intensive industry that benefits from low rates of interest, low rates of foreign exchange, and cheap prices for domestic inputs. Industrial labor unionizes and is granted minimum-wage rates that, despite erosion in real terms during inflation, draw rural labor into urban unemployment. Basic amenities for the urban population impose a growing burden on the government budget and reduce government's contribution to the flow of savings.

Fiscal policy in this context is easy to predict. By one device or another, including marketing boards, substantial revenues are collected from traditional exports. Excises on some items of luxury consumption and duties on imports of goods that compete with products of capital-intensive industry are other ranking revenue sources. Real property is taxed lightly, if at all; various tax concessions are arranged for the industrial enclave; and public enterprise supplies cheaply, or at a loss, various utilities that, like savings and foreign exchange, are rationed in some degree of compliance with a plan of industrialization. Total revenues tend to be inelastic to both inflation and growth in real national output.

Excess demand for savings spills into the capital account of the balance of international payments. In the worst of circumstances, it draws on high-cost, short-term credits until some solution is found for funding. It may be satisfied by direct private investment from abroad and by grants of official aid, some of them with local-currency counterparts that provide a measure of relief for the fiscal budget.

This is a strategy of transformation and structural change. It manipulates a congeries of relative prices—for savings, foreign exchange, basic materials, labor, government services—and it resolves by rationing the excess demands that emerge at these prices. The general purpose, of course, is to break away from relatively low levels of income and consumption by changing the national matrix of products and their inputs. It is an inferior strategy under the best of circumstances and a self-defeating strategy in the usual case: the excess demands or "gaps" that it generates on some markets and the excess supplies that occur especially in markets for labor and for products of capital-intensive industry prove impossible to close
at desired or acceptable growth rates of income and consumption. They repress development.

Why countries adopt this strategy of excess demands is not clear. Some of them seem to slip into it by inadvertence. An initial disturbance, such as a turn in the terms of external trade against traditional exports, may generate shortages of savings and foreign exchange. If one response is an increase in the rate of inflation, to cover a government deficit or to finance the bill of imports, and if appropriate changes are not tolerated in relative prices for savings and foreign exchange or for new exports and government services, the process of repression can be off to a fast start.

In some instances the strategy of interventionism with fixed nominal prices and rationing on some critical markets seems to be a deliberate choice. "Market forces" are mistrusted on the grounds that elasticities of response to relative prices are thought to be too high or too low for desired outcomes, that markets are vulnerable to exploitation or that "this country is different." Although markets are mistrusted, there is faith in the capacity of the civil service—especially if it includes a planning commission—to identify and establish the economy's appropriate growth path.

The choice of strategy may be a reflex to experience with colonialism in one form or another. The old regime may have left a relatively strong governmental apparatus, a relatively weak stratum of endogenous entrepreneurship, and comparatively primitive domestic markets, especially for finance. Costs of learning the development process by doing it in a context of effective markets and flexible prices can appear to be unacceptably high, and the quick way to postcolonial economic independence may seem to be the interventionist way.

Class interests are involved in the choice of development strategy. An entrenched civil service, which can make way for recruits from a rapidly growing labor force, may be reluctant to step aside for "market forces." Groups that benefit from the monopoly rents of import licenses, credit and investment permits, tax concessions and cheap supplies of goods and services from government enterprise, urban concentration around subsidized industrial enclaves, and shifts in terms of trade against domestic agriculture have a stake in the excess-demands regime.

Economic myopia may be partly to blame. Possibly it is overlooked that negative real rates of interest that make savings cheap for investors also make savings scarce or that overvalued foreign exchange rates that make foreign currency cheap for importers also make production and trade unattractive to exporters. Sometimes the point is missed that a government pledge to reduce conventional taxes may commit the government to the inflation tax or, possibly, to expensive borrowing abroad. Microeconomic decisions to benefit one segment of the economy may be taken without due concern for their impact elsewhere.

Distribution Function

Because the price of a product is the sum of the costs of inputs that produce the output, prices influence the distribution of income. The costs are wages to labor, rent to landowners, interest to owners of capital, and profits to entrepreneurs. Changes in income distribution among the owners of these resources are therefore determined by changes in prices. This distribution is the functional distribution of income.

Beyond the functional distribution, however, the distribution of income among households or income groups will be determined by prices in other ways. One way is through the choice of technique in producing a given output—for example, whether to use a higher or lower labor-capital ratio to produce a given output. The cost of labor relative to the cost of capital equipment will determine the choice of technique when there is some substitutability between labor and equipment. A more labor-intensive technique will clearly aid the distribution of income in favor of labor.

Another income distribution effect is related to the composition and volume of total output as determined by the pattern of demand. The impact on agricultural output is especially important because a large proportion of the population is in the rural sector, and the majority of those in absolute poverty target groups are small landowners and landless laborers.
Further, the influence of government expenditure on prices—especially through the use of taxes and subsidies—will have an effect on different income groups as owners of factors of production and as consumers of goods and services.

These various distributional implications of pricing policy are examined in the following readings. Selection 2.10 illustrates some distributional effects of transport. Selections 2.11, 2.12, and 2.13 refer to distributional implications of agricultural policy. Finally, Selection 2.14 considers the pricing of services received from the government. The latter is especially relevant for the "basic needs" approach to poverty alleviation. The fulfillment of minimum human needs requires the provision of housing, water, sewerage, education, and health. When these services are provided by government, it is important to determine how the fiscal budget—through subsidies and public pricing practices—directs the allocation of resources to meet basic needs. As indicated in Selection 2.14, the role of the government in the basic needs strategy will be all the more important the smaller the trickle-down effect of GNP growth on the income of the poor; the smaller the marginal propensity of the poor to spend on basic needs; and the smaller the possibility of the private sector supplying these services because of complementarities and economies of scale in these sectors.

2.10 Implications of Factor Choice and Technological Change in Transport

Two complex and interrelated aspects of the current transport scene have important policy implications. One concerns the choice of technology that developing countries can exercise in improving their internal transport systems. The other involves the implications of the technical improvements and innovations in international transport being introduced into these countries from developed countries. These issues arise because of the relative scarcities of capital and skilled labor and the abundance of unskilled labor in some developing countries. Differences in factor proportions as compared with developed countries have led to proposals for a greater use of labor-intensive techniques to create employment and thereby ease political and social problems. Some economists argue such techniques will improve the allocation of economic resources, relieve the foreign exchange constraint on development, and improve income distribution.

Savings in foreign exchange from labor-intensive techniques would result in two ways. First, as one commentator has observed, "the big waste of capital in underdeveloped countries has come mainly in substituting capital for labor in moving things about . . . The bulldozer, the conveyor belt, and the crane usually achieve nothing that labor could not do equally well. They spend foreign exchange solely in order to produce unemployment." Labor-intensive techniques would avoid the need to import such capital goods and spare parts. Second, labor-intensive techniques would put purchasing power into the hands of people with greater propensity to spend on domestically produced goods and services than on imports and thus ease the foreign exchange situation. The employment generated in this process would stimulate domestic industries and, by reducing unit costs and prices, have a general multiplier effect on the economy. Expansion of these industries is more likely "to raise the economic well being of the mass of the people [which] still appears to be more appropriate than gross output regardless of type or distribution."

Modern techniques of constructing transport facilities make substantial use of mechanical equipment and power tools for digging, earth moving, tamping and compaction, scraping, rolling, hauling, and other


purposes. Loading and unloading are increasingly done by forklift trucks, cranes, conveyor belts, and powered tractors. Historically, major civil engineering works in transport have been built with an extensive use of labor and relatively simple tools, as the railways and ports in developed and developing countries show. Labor could be, and has been, used in place of mechanical handling equipment, except for very heavy indivisible loads such as locomotives, electrical generators, boilers, and tanks (civil and military), which, for example, might require lifting from ship to shore. The main opportunities to use more labor and less capital in the transport sector, therefore, would seem to be in construction and handling activities. There are fewer opportunities in varying crew sizes, for example, in the operation of trucks, trains, and aircraft. Variation in the crew sizes of ships may offer more scope but may be offset by the frequently ignored foreign exchange costs of crew expenditures overseas and the costs of moving ship crews by air from port to port to pick up ships. Opportunities may also exist by varying the degree of mechanization in maintenance and administrative operations. The possibilities outside construction and loading, however, appear unlikely to have a serious effect on the basic labor-surplus problem and are of questionable value.

The apparent opportunities, and the evidence of historical experience, are strong a priori arguments. The costs of doing without or minimizing mechanical help have to be carefully compared, however, with losses in product quality and time that may occur. While “hand-made” may be synonymous with quality for some products, it does not necessarily apply in transport. Effective compaction of roads, rail track, and airport runway beds, for example, usually cannot be achieved without machines or added maintenance costs. The mechanical batch-mixing of concrete provides a uniformity of quality unattainable by hand-mixing and may be vital for safety purposes in some structural works. Rock cutting to any significant depth is generally more effective with machines. Signaling and traffic control systems that minimize the element of human error tend to be more reliable and safer.

Inefficient choices of modes, construction techniques, and equipment use may result from institutional factors that prevent or limit the pricing mechanism to reflect the real costs and scarcities of factor inputs. Overvalued exchange rates effectively underprice the real cost of imported equipment. Capital (including foreign exchange) may be made available to public sector transport enterprises at low interest rates which do not reflect its opportunity costs or make allowance for risk and inflation. By contrast, private sector trucking firms may have to acquire equipment at “true” foreign exchange rates (for example, the Pakistan bonus-voucher scheme) and to borrow at market rates of interest from private finance houses or other sources. Tax policies may also affect equipment choice or construction techniques. Finally, market wage rates may not reflect the relative abundance and low opportunity cost of unskilled labor. The conjunction of all these factors often leads investment decisions toward capital- rather than labor-intensive techniques and wrong modal choices. Experienced practitioners of investment appraisal have pointed out, however, that shadow prices for labor “should be used with caution and only for projects where the proportion of skilled or unskilled labor is unusually large... In most cases, the use of shadow wages can be safely omitted.”

The theoretical validity of the case for permitting the price mechanism to do a better job in directing demand to get investment signals right is clear. Centrally planned economies and some market economies use shadow or accounting prices as proxies or supplements to the pricing mechanism for this reason. Some developing countries have modified the gross distortions by making state-owned transport enterprises pay for equipment of imports in local currency at rates more closely approximating the “real” foreign exchange rate. The nature of the problem, however, needs to be more widely understood by political leaders. New incentives are also required to induce planners, managers, and contractors to consider labor-intensive techniques. More trained staff will be needed in most developing countries to monitor and adjust shadow pricing as a basis for investment planning or the pricing of transport outputs. Advice will also need to be more firmly underpinned by research relating to specific processes, techniques, and cases rather than generalized argument. Finally, since “a policy of dismantling controls, including [amending] labor codes, is generally unacceptable” in many of these countries, alternative project evaluation techniques will be required. These steps will take time. While “this is not much help to policy makers, for all it says is that policy must vary with circumstances,” a pragmatic approach is more likely to be of success than declarations of dogma.

Labor-intensive projects often take longer to complete. A Japanese National Railway study showed that rails could be renewed four to five times faster by a rail replacement machine than by manual labor. There is also a limit to the number of people who can be effectively employed on one construction site without impeding progress or creating serious logistical problems of food supply and housing as well as social problems if a large labor force has to move as

construction work progresses. Extended construction periods have numerous costs. Work partly completed may be damaged or remain unprotected from weather and other factors. Capital equipment associated with a project—it is unlikely to be 100 percent labor intensive—may be underutilized because of scheduling difficulties or the costs of moving it to and from the site. By delaying the appearance of expected project benefits, lengthy construction periods may, paradoxically, affect the economic viability of the project if high discount rates are used to convert future benefits to present values.

Some economists have advocated a deliberate though not generalized policy of labor-intensive, or capital-shallowing, techniques. Others have pointed out that most “practical suggestions have been limited to simple public works and farm implements but public works to make work may not be very productive”; that “interest should not be in employment only as an end in itself... [unless] there is no way of creating a more worthwhile utilization of human resources”; and that rather than a concentration of labor-intensive techniques—useful as they may be—what is required is a deliberate strategy to increase employment generally by directing demand to construction (especially house building) and mass consumption industries. Such a strategy, accompanied by a deliberately associated foreign exchange policy, it is claimed, would mean that “there would be no good reason for not utilizing highly capital intensive processes, if highly physically and value-productive, to break... bottle necks in... transport.”

In other words, the transport sector seems unlikely to be a major element in the solution of the general problem other than in parts of infrastructure construction.

Inappropriate choices of mode, uneconomic choices of technology within specific modes, and wrong factor proportions in the construction of infrastructure have occurred. Railways have been modernized and expanded when highway transport would have been a more appropriate choice, and vice versa. Roads have been built in techniques and equipment that are afterwards seen as capital intensive, lumpy in investment size, and labor saving and which were developed for use in high-wage, capital-abundant countries. Railways have mechanized operations such as track laying and track maintenance, which appear to offer opportunities for labor-intensive techniques.

2.11 Agricultural Pricing Policies and Income Distribution

Low food and other agricultural prices are politically popular on the grounds that they increase real incomes and employment of the urban poor and that the only losers are large farmers. There is a high cost attached to such low prices, however, if their effect is to retard cost-reducing investment and innovation in agriculture. Competition between farmers and traders ensures that most of the decline in unit costs will be passed on to consumers in lower prices, though per capita farm income will also rise as output rises. Moreover, studies of the effects of price distortions and controls in specific countries usually conclude that it is the upper- and middle-income urban groups (including employers) and large farmers who are the chief beneficiaries.

Arguments that higher food prices hurt low-income (particularly urban) consumers who have to buy most of their foodstuffs are usually based on the short-term income effects of higher food prices on the assumption of unchanged incomes (i.e., higher relative prices). These studies tend to ignore even relatively short-run adjustment processes that reduce the income loss, such as shifts in consumer demand toward substitute foodstuffs, and the fact that in low-income countries urban wages usually respond fairly quickly to the cost of basic foodstuffs.

Given the relationship between wages and food prices, it seems probable that food prices may have more effect on urban profits than upon real incomes of urban labor. The logic of the common argument in developing countries that food prices must be kept low in order to ensure the competitiveness of manufactured exports rests entirely on this assumption. The benefit to employers of below-market agricultural prices is even clearer in the case of nonfoodstuffs. Cotton yarn and textile manufacturers in Pakistan, for example, for many years have been able to buy cotton from farmers (many of whom are quite poor) at no more than two-thirds and sometimes half of world market prices because of foreign exchange and export tax controls, but they sell much of their output at world market prices.

Price and production controls nominally intended to provide low-cost food to poor urban groups also tend to divert production away from those crops. In Peru, low official prices have greatly reduced pro-

duction of frijol canario, a popular bean which has been a major source of protein for low-income urban consumers, and the limited output is being channeled through black markets at prices about 60 percent above the official control prices. In Egypt, farmers have increasingly diverted land, fertilizer, and other inputs from growing wheat, maize, rice, and cotton—for which they are given quotas to sell to the government at low, controlled prices—to growing fruits, vegetables, and livestock, which are not price controlled. This diversion helps ensure supply of these latter foods for middle- and upper-income urbanites, though at a high cost in terms of the balance of payments (lower export earnings from cotton and rising wheat import costs), and in massive budget subsidies to lower the retail price of imported wheat.

Food price controls also frequently benefit middle- and upper-income urban groups at the expense of lower-income rural producers. In Kenya, for example, price controls on meat and maize transfer income from low-income herdsmen and farmers for the benefit of middle- and upper-income urban dwellers. In Peru, middle- and upper-income urban groups are the major beneficiaries of large government expenditures to subsidize the retail prices of imported wheat and meat. This subsidized import of wheat has also lowered the incomes of the nation's poorest farmers—those in the high-altitude Sierra region—who have thereby lost most of their cash markets for wheat and wheat substitutes such as quinoa and potatoes.

The fact that national food subsidy schemes funnel food largely into urban areas hurts the often poorest large segment of society, namely landless laborers in food-deficit areas. Movement of significant amounts of grain in Kenya, and until very recently in India, for example, have required a government permit. The result has been to depress prices in surplus food-grain production areas but to raise them in deficit rural areas. The grain collected in surplus areas is sold in Nairobi, Delhi, and other centers at controlled prices, with significant portions going to government employees and other relatively well-to-do groups (access to the ration system for the very poor and the illiterate is often difficult). Because of the official preemption at harvest time of most rail and other freight facilities for the transport of grain into urban areas, grain prices in chronically deficit areas of Kenya and India, even at peak marketing seasons, have not uncommonly been as much as twice the government procurement price. Consequently, landless laborers and other net purchasers of food in these generally lowest-income rural deficit areas must pay much more for their food than the higher-income urban purchasers of government-subsidized food grains, or the relatively higher-income landless laborers in surplus production areas such as the Punjab.

Such sharp regional price variations also adversely affect national production. While farmers can get high prices for expanding production of the food grain in which their area is deficient, these are normally high-cost and inefficient areas for such additional production. Indian farmers in Hyderabad, for example, try to grow wheat for which they have inadequate rainfall and must use irrigation water that (either at the national procurement price for wheat or at world market prices) could be more profitably used for other crops.

**Marketing Controls**

Most developing-country controls on price margins and marketing between farmers and consumers are supposed to protect consumers against "monopolistic" traders. The few significant studies of such controls, however, suggest that such interventions frequently have very adverse effects on efficiency, production, and income distribution. The most comprehensive Indian study concluded that the private grain market was highly competitive, that traders operated efficiently within government-imposed technological and policy constraints, and that government efforts, rather than being expended in market controls, could have been much more fruitfully directed toward helping improve the competitive environment, including improving farm-to-market transportation facilities, standardizing weights and measures, grading produce, disseminating market price and stock information, and managing national support price and buffer stock programs. Periodic threats by the Indian government during the prior twenty-five years to nationalize all trade in grains, and the issuing of new rice-milling licenses primarily to cooperatives and public sector firms, however, had resulted in very little private investment to improve either storage or milling facilities. Furthermore, government-agency marketing costs are several times as high as those of private traders, who flourish despite government policies.

A study in Peru concluded that marketing measures initiated to eliminate abuses by middlemen represented one of the major constraints limiting agricultural production and might lead to destruction of the pool of private marketing talent with no comparable gain in public sector expertise. Because of stringent laws and penalties against speculation and monopolization, there is little private investment in storage facilities in Peru. This lack of storage facilities results in periodic gluts, scarcities, and excessive price fluctuations as wholesale truckers move products between consumption centers on a day-to-day basis to attempt to take advantage of spatial price differences. Lack of accurate market information is another problem in Peru. A large proportion of transactions takes place on the black market, but only official prices
are reluctant to provide accurate price information for fear of fines and jail sentences for price violations.

**Input Subsidies and Income Distribution**

Farm subsidies also benefit primarily middle- and upper-income farmers in low-income countries. Large-scale farmers buy most subsidized inputs. Poorer farmers usually lack the money to buy adequate amounts of fertilizer and pesticides, and are commonly unable to get credit except at near-prohibitive private rates of often 60 to 100 percent per year. Even in countries with subsidized bank credit for agriculture, rich farmers get most of the credit because of legal or administrative restrictions and through open or disguised bribery. Credit and subsidy programs for tractors, tube wells, and other fixed investments also go mostly to the largest and richest farmers. Data from Kenya indicate that about 80 percent of fertilizer, which has been subsidized for many years, has gone to large farmers. Also, farmers with fewer than forty acres spend no more than one-fifth as much on tractor cultivation as do the least mechanized of the large farmers. This serves to skew the farm income distribution by reducing agricultural employment and concentrating farm income. Yet, despite the high foreign exchange and government budget cost of large-scale mechanized agriculture, Kenyans farming fewer than ten acres produce substantially more maize and employ much more labor per acre than do large farmers. While comparable figures are not available for most other poor countries, observation supports the same kind of skewedness in the use of purchased inputs, with similar consequences.

Water is also a subsidized input, provided at less than its cost to the government, in most countries with large public irrigation systems. The farmers who receive this subsidized water generally have substantially higher incomes (because of the water) than farmers without access to public irrigation. Thus, claims that water should be subsidized to help small farmers without access to public irrigation. However, the largest farmers often obtain more than their normal entitlement by bribes to the local authorities.

2.12 Agricultural Prices, Distribution, and Employment

The simplest policy issues with respect to agricultural prices are those arising from the question: If all other influences are constant, what is the distributional effect of a change in relative agricultural prices on the absolute and relative levels of income of various consumer income classes? But this question has very limited relevance. The most frequent cause of change in relative agricultural prices is a shift in supply, to which consumption must be adjusted. Hence the extent of an increase in price and its consequent distributional effect is also a reflection of the related changes in incomes of both producers and consumers, and the consequent interacting demand and supply effects. Moreover, change in relative price may influence the level of production in a future period, and perhaps less by movement along a production function than by shifts in the production functions, so causing later, opposite structural effects on relative prices. The fact that agricultural commodities are the principal wage good for low-income families creates a further interaction with employment and income questions. The term “agricultural commodities” embraces not only food-grain but also a wide range of non-food-grain products. Within the context of a static technology, aggregate production of food-grain tends to be relatively inelastic with respect to price while aggregate demand, at least in relation to other commodities, is also relatively inelastic with respect to price. But for much of the non-food-grain category, supply and demand are relatively more price elastic and production can frequently be highly labor intensive. Thus changes in relative prices and in the production of non-food-grain agricultural commodities not only play an important role in determining employment and income levels of the poor, but are themselves very much a function of the underlying structure and dynamics of income and income distribution.

**Distribution of Income among Consumers**

Change in food-grain prices causes a larger percentage change in the real incomes of low-income consumers, but a larger absolute change in the real incomes of high-income consumers. The absolute effect on income of high-income consumers may have secondary effects on the poor through changes in consumption of other goods and services and consequent change in employment in their production. For example, in India the top 5 percent of the income distribution spends more than two and a half times as much per capita on food grains as the lowest two deciles in the income distribution. Despite its large absolute expenditure, this upper-income class allocates only 15 percent of its total expenditure to food grains, compared with 54 percent for the lower-income class.

With the use of simplifying assumptions, an attempt is made to show the implications of these relations. First, only the income effect of an increase in food-grain price is examined. For high-income
nations this effect is normally considered negligible, but it is a major factor in low-income nations, where the proportion of family income spent on food grains is much higher. The following analysis uses Indian data because of their availability, because food grains are dominant in Indian consumption patterns, and because of the low levels of income. After the income effect is analyzed, alternative assumptions with respect to substitution effects will be discussed.

The income effect of a price change is analyzed by using cross-section data to compare the initial expenditure patterns with those of the different total expenditure class entered as a result of change in real income owing to change in food-grain prices. For the data presented, if food-grain prices rise by 10 percent, the lowest two deciles in expenditure, which initially spend Rs4.830 per capita per month on food grains, experience a real total expenditure decline of Rs0.494 per capita per month. The decline is in fact slightly greater than 10 percent of the initial expenditure on food grains because the proportion of total income spent on food grains increases in successively lower income classes. In this case, if all other prices, as well as the proportion of income saved, are assumed to be constant and expressed in constant prices, the new real expenditure pattern will be Rs8.436 per capita per month, or a decline of 5.5 percent.

In contrast, the top 5 percent in the expenditure distribution experiences a decline in real total expenditure of Rs1.01, or somewhat more than twice as large an absolute decline as that of the lower-income class—but at only 1.2 percent of the initial expenditure, less than one-quarter as large a percentage decline as that experienced by the lower-income class. The expenditures of the poor on food grains are, of course, far more elastic with respect to price than are those of the rich. For the bottom two deciles in the income distribution, the elasticity of response of real expenditure (assuming no substitution effect) to change in food-grain prices is 0.55, compared with 0.12 for the top 5 percent.

Other relations important to income distribution are illuminated by this exercise. It is significant that in response to a price increase, both the absolute and the percentage decline in expenditure on food grains is greater, the lower the income class. Thus for a 10 percent increase in food-grain prices, the bottom two deciles reduce their real expenditure on food grains by 5.9 percent, compared with a reduction of only 0.2 percent by the upper half of the tenth decile. More important, this top income class reduces its absolute real expenditure on food grains by only Rs0.026 compared with more than ten times as large an adjustment by the lower-income class. These data illustrate that in a market economy the bulk of adjustment to reduced supplies of food grains is made by low-income consumers. Given the low initial level of food-grain consumption in the lower-income deciles, the privation imposed on them by rising grain prices is very great. But any measures which seek to insulate the poor from the necessity to adjust to reduced supplies will force the need for adjustment on to those whose demand is much more inelastic—thereby causing proportionately much greater, and perhaps explosive, price increases. It is clear from this analysis that it is essentially impossible to protect the poor from the principal income effects of a short crop by market measures.

Further adverse consequences for the poor of a change in relative food-grain prices may follow from the effects of such a change on the consumption of other commodities. Consequent to a rise in food-grain prices, the absolute decline in expenditure for almost all non-food-grain commodities is greater for higher-income classes than for lower-income classes, although of course the percentage decline in expenditure is much greater in each case for the lower-income classes. For example, in response to a 10 percent increase in food-grain prices, the lowest two deciles in the income distribution reduce consumption of milk and milk products by 18 percent, compared with a reduction of 1.3 percent by the top 5 percent, though the absolute reduction by the higher-income class is more than three times as great as for the lower-income class. Overall, while food-grain consumption declines only 1.5 percent, consumption of most other commodities declines between 4 and 8 percent.

These comparisons suggest that an increase in food-grain prices may lead to substantially reduced consumption by the poor of agricultural commodities of high nutritive value. It is of course conceivable, but neither logical nor likely, that the poor would respond to increased food-grain prices by some substitution of higher-quality foods, which, though more expensive, had not increased in price.

In addition, the large absolute reduction in consumption by higher-income groups of goods and services other than food grains, and particularly of livestock products and vegetables, the production of which is generally highly labor intensive, reduces employment. To the extent that such reduction in employment reduces incomes and hence demand by the poor, the increase in price of food grains will be dampened. This indirect effect of food-grain price changes of course complicates empirical analysis: the greater the employment effect, the less will be the price increase. But either way the poor pay—in terms of lower real wages as prices rise, or in reduced employment caused by the decline in real income of the upper-income classes.

Distribution of Income among Producers

The effects of relative price changes on agricultural producers differ from the effects on consumers in two
important respects. First, the income effect, assuming production is constant, is in the same rather than in the opposite direction as the price change. Second, the largest effects, both relative and absolute, fall on the higher-income producers with the largest marketings.

The effect of a price change that occurs independently of change in the volume of domestic production is easier to analyze than the effect of price changes in response to production changes, perhaps induced by variation in weather. These two somewhat different cases are discussed below and are followed by an analysis of the effects of price changes on production.

EFFECT OF CHANGE IN RELATIVE PRICES WITH PRODUCTION CONSTANT

The relations described below refer to effects of price changes on producers when quantity produced stays constant, as could happen if the price changes were owing to a foreign food aid program. It could also happen as a result of commercial trade, although in that case the relations are more complex because change in imports and exports of food grains would presumably be offset by trade in other commodities, with further price and income effects.

The effect of relative change in agricultural prices on producers' incomes depends on the quantity they produce, the quantity of home consumption and hence of marketings, and the quantity of purchased production inputs. The effect of price changes is much greater in both absolute and percentage terms on larger farmers than on smaller farmers—the larger farmers normally produce more, market a higher proportion of their production, and have a higher proportion of output represented by purchased inputs.

Thus, if a food-grain price increase is seen as a simple transfer of income from consumers to producers, it largely takes place between high-income consumers, who expend the largest absolute amount on food, to high-income producers, who market the largest absolute quantities. In terms of percentage change in income, the transfer causes the largest decline in the income of low-income consumers and the largest increase in the income of high-income producers.

Effect of Relative Changes in Agricultural Prices on Agricultural Production

The rate of increase in food production can be a major limitation on the rate of increase in employment. Without a commensurate increase in the supply of food, growth in the number of those in paid employment and the consequent rise in demand for wage goods, of which food constitutes a dominant share for low-income workers, will cause an increase in food prices, the effects of which may substantially neutralize the benefits from increased employment. Adequate additional supplies of such wage goods can rarely be mobilized through reallocation of existing domestic resources or through international trade, and policies on employment, agricultural production, and agricultural prices must thus be closely linked. Change in relative prices plays at most a very limited role in increasing agricultural production in the context of traditional technology. But if used to complement technological change in agriculture, price policy may speed the growth of production significantly. In this context, too, price policy may encourage the adoption of innovations by low-income producers through its influence on profitability as well as on risk and uncertainty.

2.13 Consequences of Farm Tractors in Pakistan

This study examines the major consequences of the introduction of large-scale tractor technology to farms in Pakistan. The assessment is based on the use and effects of tractors financed by World Bank lending. This comprised three International Development Association (IDA) credits to the government of Pakistan, distributed through the Agricultural Development Bank of Pakistan (ADBP) beginning in the latter part of the 1960s. The study was initiated after the approval of the third credit. It arose, inter alia, from a growing concern about possible adverse social effects of farm mechanization, particularly the effect of tractors on farm employment.

Survey Data

The assessment is based on information obtained from a survey of some 200 farms randomly selected from the 3,868 that received tractors from the first credit in 1967–68. To evaluate the effect of tractors on the recipients, field survey data were collected on farm structure, resource use, production processes, and output in both the “before” and “after” tractor situations.

FARM SIZE

The most striking change after the introduction of a tractor was the growth in the average size of the farms by a factor of 2.4—from 45 acres to 109 acres per farm for the survey as a whole. Only 23 of the 202
survey farms did not increase their acreage. The sources of additional land were:

<table>
<thead>
<tr>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previously uncultivated land</td>
</tr>
<tr>
<td>Previously rented out land</td>
</tr>
<tr>
<td>Newly rented in land</td>
</tr>
<tr>
<td>Newly purchased land</td>
</tr>
</tbody>
</table>

All of the changes subsequently discussed are convoluted with this growth in farm size and the related structural alterations.

CROPPING INTENSITY

The first evidence of a change in the intensity of land use is that the proportion of farms growing each crop increased. Thus the mean number of crops grown per farm shifted from 4.77 to 7.39, fairly equally spread between both kharif and rabi seasons. The overall cropping intensity also increased from 111.5 to 119.0 percent. The increase was about equal in both seasons. There is some evidence that the increase in intensity was greater on the smaller farms, suggesting that those farms which expanded less in acreage increased their cropping by intensification.

CROPPING PATTERNS

The area sown to each crop increased in absolute terms, in keeping with the farm size change and increased cropping intensity. The exception was a decrease of 50 percent in the area devoted to fodder crops. The land so released was allocated mainly to wheat in the rabi season and rice in the kharif. There is little evidence of a concerted shift toward higher value crops but there was an increase in vegetable and other minor crops (from 2.3 to 6.0 percent of cropped acreage) notably on the larger farms. There is no convincing evidence of beneficial changes in crop yields associated with tractor use. The increases that were revealed can be attributed largely to increased fertilizer use on high-yielding varieties.

LABOR

The labor used per farm increased overall for every size group. Allowing for farm size changes, however, the labor use per cultivated acre decreased by some 40 percent. This was exacerbated by tenant displacements, which averaged 2.25 per farm from the original acreage farmed plus a projected 1.95 per farm from the area purchased or rented (assuming the same ratio of one tenant to 11.2 acres)—totaling 4.2 tenant families per farm. Under different assumptions as to the amount of labor provided by tenant families, it was estimated that 11.8, 9.7, or 7.5 full-time jobs were replaced by each tractor. Some of these jobs were compensated for by casual labor use on seasonal tasks at a rate that implies an overall net attrition of jobs of about five per tractor.

MACHINERY PACKAGE

The tractors on the farms surveyed comprised 118 units of 45 horsepower and 90 of 55 horsepower. Overall, these were used intensively—more than 1,200 hours per year. The level of use was higher on larger farms, but smaller farms hired out more extensively. The intensity of use per cropped acre was markedly lower on larger farms. Some two-thirds of all tractor time was devoted to "cultivation"—6.25 hours per acre on the smallest farms and 4.29 hours per acre on the largest. On average, 12 percent of the total hours of use were for off-farm "transport"—varying from 18 percent on the smallest farms to 7 percent on the largest. Sixty percent of the farms in the survey had tube wells, but these had little apparent effect on major variables.

FINANCIAL RETURN

The return on investment to the individual mechanizing farmer was estimated to have been about 57 percent. This was calculated for the total area of all farms using tractors, rather than as an average for individual farms. The benefits included additional revenue from purchased and reclaimed land, multiple cropping, cash crops in place of bullock fodder, and higher value crops, as well as savings in labor wages, increased livestock output, and fees from tractor hire.

ECONOMIC RETURN

The social rate of return, similarly calculated on the "project area" basis, was estimated at 24 percent. This figure was arrived at by correcting for taxes and subsidies, omitting transfer payments such as land rental, and allowing for the substantial foreign exchange component in the capital investment and the undervaluing of domestic production by using shadow prices. The savings in wages were counted as zero.

STRUCTURAL EFFECTS

Although these rates of return on investment appear acceptable, significant adjustments remain unaccounted for in the pattern of resource use and structure of agriculture consequent upon the introduction of tractors. Were social accounting methods available for assessing the changes in wage employment and tenancy patterns, the economic benefits would be substantially reduced.

Potential Benefits of Mechanization

In the many studies dealing with the impact of tractors and tube wells on Pakistan agriculture, there
have been various estimates of the net benefits that accrue both to the mechanizing farmer and to society at large. For a variety of reasons, it seems clear that private benefits exceed social benefits, and in all studies, the computed net private benefit is such as to provide a strong incentive to the farmer to adopt mechanical technology. Determination of the net social benefits is more hazardous, and conclusions vary from the optimistic to the dismal.

Herein lies a paradox. If the adoption of machine technology can be of value to society, the private benefit to the farmer needs to be held at a high level to induce the adoption of the new machines and methods. But if the policies that provide these incentives also result in mechanization being put into effect in such a way as to subsequently lead to a net loss to society, then either the technology should be rejected or—more desirable though no less easy—a policy framework should be operated that directs the adoption and exploitation of the technology along more favorable lines. Such an approach should be possible if technological advance is, by definition, always potentially beneficial to society.

PRIVATE BENEFITS

Given the policy measures that have been in force in the Pakistan economy, all calculations seem to show that the adoption of the tractor alone, the tube well alone, or (more emphatically) a combined tube well–tractor package can show major improvement in farm profit. These results, which have been derived in different studies by a variety of means from single predictive budgets, complex programming models, and examination of the accounts of mechanized farms, seem dependent upon two facts. First, there is a very real benefit to farmers, especially on large farms where the full capacity of the indivisible mechanization inputs can be effectively utilized, in terms of the lowered production costs and increased output that can follow from the adoption of the technology package. This effect is magnified in those cases in which the mechanizing farmer takes over the land formerly leased out to his tenants in order to utilize fully the capacity of his machinery. In these instances, he receives both his “normal” landlord’s share of production increases plus the tenant’s entire share. This additional benefit in itself has provided a powerful motive for the adoption of mechanization. Second, this benefit has been exaggerated in Pakistan (especially during the period of the mid- and late-1960s when much mechanization was proceeding under the IDA credits) by the exceptionally favorable factor and product prices that were maintained by the government. Thus, the prices of many crops were pegged at well above world market values, particularly sugarcane, wheat, maize, and rice.

Furthermore, not only was credit for mechanization investment made available by the ADBP at artificially low interest rates (roughly 6–7 percent compared with 12–15 percent commonly charged by commercial banks), but also the foreign exchange needed to purchase the machinery imports was made available at an official exchange rate of Rs4.75 to the dollar, whereas its real value in the economy was widely held to be about twice this. Neither was the importation of tractors encumbered by the duty and sales taxes that were imposed on other capital goods. Consequently, tractors that were imported under license at about Rs14,000 immediately had a resale value on the open market of nearer Rs25,000, reflecting their true value in production. Bose and Clark estimate that the Pakistani farmer paid about $65 per tractor horsepower compared with the $100 or more paid by his U.S. counterpart.9

Taking all these considerations into account, it is not surprising that the private rates of return on mechanization investment appear quite high. Bose and Clark estimate the private cost per horsepower at Rs200 and private return at about Rs250. The net return for the farmer adopting a 40-horsepower tractor is thus about Rs2,000 annually. The appraisal report of the first IDA credit calculated that the gross value of both crop and livestock production per acre would increase by a factor of between three and four, while costs would rise by less than three times. The conclusion is that total net return per acre cultivated will increase by more than 400 percent, giving an additional net return of some Rs425 per acre, or a financial rate of return on the investment of about 30–40 percent. Ahmad quotes a private rate of return for investment in tractors alone on a 75-acre farm at 30 percent, and for a combined tube well–tractor package, 45–50 percent.10 On all this evidence, the policies operating in Pakistan were exceedingly generous to the farmer, and presumably many of these private gains could have been taxed away or considerably reduced in other ways without having an undesirable effect on the rate of investment in new technology. Indeed, given the widely quoted divergence between private and social net benefits, perhaps such a move would have had overwhelming benefits for society by deterring mechanization in those situations in which its private profitability was more marginal.

SOCIAL BENEFITS

Because the domestic market price of crop products has been held above their value on world markets, the additional output deriving from mechanization is worth more to the individual farmer than to society. In addition, the social costs associated with mechanization are higher than those accruing to the individual. This is not only because of the subsidy element in the pricing of such inputs as fertilizers and the mechanization inputs themselves, but also because there are various societal costs that the individual farmer does not have to meet and therefore does not include in his assessment computations. For example, any net labor displacement resulting from the adoption of mechanical technology appears as a saving to the individual. To a society in which labor is not in short supply, the real saving will be less than the wage rate, will possibly be zero if employment opportunities are not freely available in the economy, and may well be strongly negative—especially if physical resettlement and retraining of the displaced labor is necessary. Added to this must be the consideration that the productive value of imported capital goods is far higher in the nonagricultural sector of the economy. There are also wider implications for domestic agricultural markets. These include the possible rapid rise in the price of meat as a result of the decline in the supply of bullock culls as a source of beef, as well as a depression in the domestic prices of crop products brought about by an expanded output from the larger, mechanized farms, both of which would have adverse effects on the incomes of the bulk of the farms that are small family holdings.

The re-evaluation of the consequences of mechanization within a social accounting framework, while involving much guesswork and assumption concerning fairly unmeasurable quantities, nonetheless seems to suggest that the net benefits to society during the short term are possibly quite low, and may even be negative. Whether these can be regarded as merely short-term social costs in the wider and longer run process of restructuring the Pakistan economy, in line with the "normal" progression in economic development, toward a more industrial-based economy with a less dominantly traditional agricultural sector, might well be a point of some contention.

Estimates of the social net returns are, by their very nature, clearly less definitive than are the assessments of private benefits. Nevertheless, of the estimates that have been made, Bose and Clark put direct social costs at Rs175 and direct social benefits at around Rs150 per horsepower.11 With the pos-

tulated indirect benefits not exceeding indirect costs, they conclude that the net social advantage from mechanization is distinctly negative. Gotsch, however, takes a much more hopeful view from his calculations and concludes that, even when valuing inputs and outputs at their social opportunity costs, the investment in tractors and associated machinery is worthwhile.12 In a similar vein, the appraisal reports of the three IDA credits were consistently expectant of a high positive social net benefit from the mechanization package, quoting economic rates of return on the investment as high as 50 percent (though on slightly different assumptions, figures of 28 percent and 12 percent were computed).

Labor and Capital Aspects of Mechanization

The potential effect of mechanization on labor employment within agriculture is conventionally viewed as having several facets:

- The replacement of bullocks by tractors results in a displacement of labor associated with the working and feeding of bullocks.
- The adoption of specifically labor-saving machinery (seeders, weeders, harvesting machinery, and so on) displaces labor from these traditional operations.
- The adoption of machinery associated with new cultural techniques (fertilizer application, chemical spraying) has no disemployment effects but may marginally raise labor use.
- The increased intensity of cropping (and perhaps yields) and the addition of reclaimed land consequent upon the introduction of tractor models should lead to the creation of employment opportunities.

A priori, the net effect of these forces upon aggregate labor employment in agriculture is indeterminate, being dependent on the relative influence of each separate factor. Proponents of mechanization predict that net labor employment will rise; antagonists argue the reverse. In the context of this survey, wherein expansion of farm size was such a dominant aspect of the adjustment to tractors, two further considerations must be included:

- The displacement of tenant and owner-occupier farmers, and their family labor, from the land that is added to the enlarged holding
- The displacement of hired labor from these farms that are absorbed.

The labor employment effects of farm mechanization are therefore complex, and careful analysis is

11. Bose and Clark, "Some Basic Consideration on Agricultural Mechanization."

required to disentangle the separate influences and their net impact. In aggregate terms, one can conceive of the mechanization process initially substituting for labor—farmers, family labor, and hired workers—followed by a partial or total re-employment, dependent on the extent to which output expansion results in new labor demands. There are, however, various more subtle effects in addition to those on the numbers of people employed. First, the farm enlargement results in the loss of "independent" farmer or tenant status by those whose farms are absorbed—even though they may be re-employed elsewhere within agriculture. Second, the characteristics of the mechanization package and its effect on output can result in a significant shift in the seasonal pattern of employment opportunities; the displacement of full-time (family or hired) labor may be nullified to a greater or lesser extent by an increased demand for casual labor, but this does not necessarily mean that the same individuals are concerned nor that the impact on employment is neutral.

In the analysis of the survey results, the treatment of labor employment on a per farm basis would have been misleading because the nature of farms changed so radically with mechanization. To draw general conclusions, therefore, changes in labor use must be related to a given area of agricultural land. In this case, labor employment on the final area of land in the survey farms (22,025 acres) would have had to be examined both before and after the mechanization program began. Unfortunately, labor use before mechanization on the 12,942 acres that were absorbed from other farms over the period could not be measured directly and therefore had to be estimated to make overall comparisons possible. A variety of alternative assumptions about the level of employment on the absorbed land area were made, and the results presented below are based on what were felt to be the most reasonable.

**TENANT AND FARMER DISPLACEMENT**

The increase in cultivated area of the survey farms was composed of:

<table>
<thead>
<tr>
<th>Land Type</th>
<th>Acres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newly purchased land</td>
<td>1,481</td>
</tr>
<tr>
<td>Newly rented in land</td>
<td>3,386</td>
</tr>
<tr>
<td>Previously rented out</td>
<td>5,228</td>
</tr>
<tr>
<td>Total</td>
<td>10,095</td>
</tr>
<tr>
<td>Newly reclaimed land</td>
<td>2,847</td>
</tr>
<tr>
<td>Total farm enlargement</td>
<td>12,942</td>
</tr>
</tbody>
</table>

The reclaimed land, being previously uncultivated, can be assumed to have supported no specific labor employment; the remaining 10,095 acres, however, were previously cultivated and their absorption into the survey farms would have meant the displacement of their tenant and owner-occupier farmers. The data recorded in the survey showed that in the appropriation of the 5,228 acres of land previously rented out, the number of tenants displaced was 456. This suggests that the mean size of tenant farm taken over was 11.5 acres. No information was available on the tenancy picture on the newly purchased and newly rented in land, but if it is assumed that it too was farmed in units of 11.5 acres, this gives an additional 423 tenants and farmers who lost their holdings. In total then, an estimated 879 tenants and farmers (and their associated labor) were dispossessed by the 202 survey farms—almost 4½ for each tractor that was introduced into agriculture. It is recognized that this is an estimate rather than a direct measurement, but given the predominantly small farm size structure of Pakistan agriculture (some 77 percent of farms were less than 12.5 acres) there is no evidence to suggest that the mean size of holding absorbed was much greater or less than the 11.5 acres used here.

**CHANGES IN FAMILY AND HIRED EMPLOYMENT**

The data recorded in the survey covered only the family labor employment situation on the survey farms as they were before the arrival of the tractor, and after its adoption; there was no possibility of directly measuring the previous labor use on those holdings that lost their identity in the farm enlargement. Consequently, estimates were again necessary to assemble the full picture of the changes that took place.

Table 2-6 illustrates the not surprising result that both family and permanent labor employment per farm increased in every size group as a result of mechanization. Despite this, labor employment per acre fell significantly, indicating that the capital investment was used to substitute for full-time labor in production. The pattern of changes in labor use is generally consistent with the pattern of changes in farm size and cropping intensity: the farms that grew least in size (group 1) showed the smallest change in labor employment per farm, and the farms that achieved the smallest gains in cropping intensity (notably group 3) demonstrated the greatest declines in the intensity of labor use per acre. All these findings are in complete accord with expectations based on the various hypotheses about labor employment that were advanced earlier.

From the sector viewpoint, the most important aspect is the effect on aggregate labor use on the land area farmed. The survey data show that total family labor increased by 139 (if part-time family workers are taken as equivalent to half full-time) and permanent workers by 294 on the survey farms, a total of 433 full-time job opportunities. These figures must be adjusted for the full-time labor displaced on the land that was absorbed into the mechanizing farms.
Table 2-6. Employment of Full-time Labor per Farm before and after Mechanization

<table>
<thead>
<tr>
<th>Size group</th>
<th>Full-time family workers</th>
<th>Total family workers</th>
<th>Permanent workers</th>
<th>Total full-time labor</th>
<th>Percent change</th>
<th>Cultivated acres per worker</th>
<th>Percent change in labor per acre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>2.18</td>
<td>2.29</td>
<td>1.88</td>
<td>4.17</td>
<td>14</td>
<td>6.33</td>
<td>-33.8</td>
</tr>
<tr>
<td>After</td>
<td>2.58</td>
<td>2.66</td>
<td>2.08</td>
<td>4.74</td>
<td>31</td>
<td>9.56</td>
<td>-44.5</td>
</tr>
<tr>
<td>Before</td>
<td>2.07</td>
<td>2.15</td>
<td>3.03</td>
<td>5.18</td>
<td>44</td>
<td>7.01</td>
<td>-50.1</td>
</tr>
<tr>
<td>After</td>
<td>2.89</td>
<td>3.05</td>
<td>3.75</td>
<td>6.80</td>
<td>86</td>
<td>12.63</td>
<td>-27.2</td>
</tr>
<tr>
<td>Before</td>
<td>1.59</td>
<td>1.65</td>
<td>4.05</td>
<td>5.70</td>
<td>13.50</td>
<td>18.54</td>
<td>-42.2</td>
</tr>
<tr>
<td>After</td>
<td>2.20</td>
<td>2.34</td>
<td>5.85</td>
<td>8.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before</td>
<td>1.45</td>
<td>1.52</td>
<td>6.54</td>
<td>8.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After</td>
<td>1.83</td>
<td>2.02</td>
<td>12.95</td>
<td>14.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All farms</td>
<td>1.93</td>
<td>2.02</td>
<td>3.34</td>
<td>5.36</td>
<td>40</td>
<td>8.40</td>
<td>-42.2</td>
</tr>
<tr>
<td></td>
<td>2.56</td>
<td>2.71</td>
<td>4.79</td>
<td>7.50</td>
<td></td>
<td>14.53</td>
<td></td>
</tr>
</tbody>
</table>

a. Part-time family workers taken as equivalent to half full-time workers.

As already estimated, 879 small farms were eliminated in this process, but no direct measure is available of the full-time labor that these farms employed. Table 2-7 indicates that farms in the smallest size group, when operated by bullock methods (that is, before mechanization), used an average of 2.29 family workers and 1.88 permanent workers. These figures, however, refer to a mean farm size of 26.4 acres, rather larger than the 11.5 acres adopted for the absorbed farms, which therefore probably had a somewhat lower labor employment. It would appear from table 2-7 that family labor is higher per farm, and permanent labor lower, as farm size decreases. This gives some basis for making alternative assumptions about the original labor employment on the land area taken over by the mechanizing farmers. Alternative estimates are given in table 2-7.

The four employment configurations suggest a range of from 3,665 to 2,417 workers displaced. Of these four alternatives, possibly the third is the most reasonable, suggesting the small farms that were absorbed employed 2.5 family workers and 0.5 permanent workers. On this basis, an estimated 2,636 full-time employment opportunities disappeared in the farm growth that accompanied mechanization. Correcting this for the recorded increase in employment on the survey farms of 433 results in a net loss of some 2,203 full-time employment opportunities brought about by the introduction of 208 tractors, an average of about 11 per tractor.

In the process, an estimated 879 farmers and tenants and their 1,318 family workers lost their independent livelihood. The full implications of this effect may not be contained merely within these figures. It may have been that many of these displaced workers found re-employment within agriculture—though in a labor surplus sector this may mean that they swelled the ranks of the "disguised unemployed." If under such circumstances the family farm is a convenient repository for the disguised unemployed, at least providing them with subsistence consumption, the family workers displaced may have been transferred directly into real unemployment.

If all these estimates are drawn together, overall employment can be computed. In the previous section it was seen that 2,203 full-time family and hired workers were displaced from the land in the survey farms; these workers were partly replaced by the equivalent of 496 casual workers, giving a net figure of 1,707 full-time employment opportunities lost as a result of the mechanization on 202 farms—well over eight jobs per tractor. This loss would seem to lend support to those who express fears that the mechanization of traditional agriculture will lead to an unwanted and significant labor substitution effect. In the context of this survey, the result is unsurprising, given the limited extent to which cropping intensity and output expanded on the land brought under mechanization. While labor substitution can-
Table 2-7. Estimated Initial Labor Employment on Farms Absorbed by Tractor Farmers

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Family workers per farm</th>
<th>Permanent workers per farm</th>
<th>Estimated displacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.29</td>
<td>1.88</td>
<td>2,013</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>1.0</td>
<td>2,197</td>
</tr>
<tr>
<td>3</td>
<td>2.5</td>
<td>0.5</td>
<td>2,197</td>
</tr>
<tr>
<td>4</td>
<td>2.75</td>
<td>0</td>
<td>2,417</td>
</tr>
</tbody>
</table>

Note: The number of farms absorbed is estimated at 879; the size of each at 11.5 acres.

not be shown to be a necessary consequence of tractor adoption, it was clearly the chosen direction of adjustment for the farmers in this survey.

Two final caveats must be added to the above estimates of the net employment effects of the mechanization program:

1. The tractors in the survey spent much of their time providing custom services to other farms. To the extent that this resulted in output expansion on these farms, there might have been an added (unmeasured) employment creation effect; but there might equally have been a labor displacement effect if these farmers replaced some of their bullocks, or displaced their tenants and increased their farm size, with the availability of hired tractor services. The net impact on other (nonsurvey) farmers is therefore undeterminable.

2. The arrival of the tractor presumably resulted in the creation of some job opportunities in the sectors ancillary to agriculture (particularly in machinery servicing and repair), which cannot be accounted for here.

Even if the most generous allowances are made for the above points, it would appear that the equivalent of at least five full-time employment opportunities in farming disappeared for each tractor introduced into Pakistan under the first IDA credit for agriculture.

2.14 Who Benefits from Government Expenditure: Colombia

To identify the specific policy options that can be used to change the composition and direction of public expenditure, the distributive direction of the present public expenditure must first be evaluated. This study makes such an evaluation with the help of a case study. Colombia was chosen because it represents a typical middle-income ($600 per capita income), semi-industrialized Latin American country. It has a growing urban population, strong income differences between regions, and both commercial and small subsistence farming. The data base is reasonably good, and earlier work on the subject allows comparison of results.

Objectives and Limitations

A full evaluation of the distributive effect of government expenditure—the change in the real income of individuals resulting from the total or marginal presence of the government—is beyond the scope of this research, since it would require a model of the determinants of the incidence of government intervention through factor and final goods markets. When this intervention provides public services whose benefits may not be completely internalized in market demands (for example, health), an additional difficulty arises. The benefits from the extra consumption of these services cannot be approximated by the prices individuals were paying for private substitutes before public intervention. Thus, tracing such benefits becomes a problem analogous to that in social project evaluation.

This study addresses a more restricted set of questions, which can be summarized as follows:

- How is the consumption of publicly provided services distributed by income groups, regions, and other relevant characteristics associated with income levels?

This study covers only services for which individual or household consumption can be identified and, if possible, measured. The typical pure public good, such as defense or justice, is excluded.

- What is the subsidy associated with the provision of these services by the public sector?

Two definitions of the subsidy are of interest. The first definition is the subsidy derived from the government pricing policy: in other words, the difference between the long-run cost of the service to the gov-
government and the price to consumers. It takes the presence of the public sector as given and simply asks about the distributive effect of the government pricing policy. The second definition is the subsidy of having a public sector with that particular pricing policy as opposed to a situation in which equivalent services would have been provided by the private sector. This subsidy consists of the earlier definition (the pricing policy subsidy) plus the difference in cost between private and public provision of the service.

For particular services, this research derives estimates for the first definition of the subsidy. Estimates for the second definition would require data on the cost of equivalent services in the absence of public intervention (for example, the cost to a household of obtaining a unit of light from oil lamps)—a task beyond the scope of this study.

- What factors explain the distribution of consumption addressed in the first question? How much can be explained by the existing structure of supply of the service and how much by factors governing the private demand for these services?

This is a difficult area to research. It is, however, the most relevant in identifying policy instruments to increase the consumption of these services by low-income groups.

To identify these policy instruments, it is useful to distinguish between two types of situations in which a particular household does not consume a public service: (1) the household does not consume the service because the supply network is not geographically present, and (2) the household is on the supply network but decides not to consume the service.

In the first instance there is no consumption because of the location of the supply network: it is basically an institutional datum determined by past investment decisions in the sector. In the second, the lack of consumption is determined by demand and results from a voluntary choice. Because the factors governing this choice (variables behind the demand) may be different from the factors determining the location of the network (variables behind the supply), it is useful to recognize explicitly this supply-demand mechanism behind consumption. This study attempts to identify the explanatory factors behind the supply and demand for particular services.

- What income groups have benefited from the changes in the supply of these services as a result of recent investment policies?

Thus far the research has addressed the distribution of consumption at one time: that is, the consumption stemming from the existing supply or stock of infrastructure. Of at least equal interest is the distribution of new consumers of the service that results from expansions of the supply over time. For those sectors in which investment has been significant as a fraction of the national budget, this study estimates the income distribution of the new households consuming the service.

Methodology and Further Simplifications

This section outlines the methodology followed to address the four questions stated above. It will become evident that further simplifications are required in addition to those already mentioned.

The subsidy received by a household from consuming a service (S) is equal to the subsidy per unit (s) multiplied by the units of that service being consumed by the household (Q). The subsidy per unit is equal to the difference between the long-run marginal cost to the government of providing the service (c) and the price charged to consumers (p). The quantity consumed (Q) is itself a function of the price charged and of other socioeconomic variables (x) influencing the household demand for the service.15

The problem of using the national budget to identify public services with significant subsidies and the limitations of using the budget to compute c, the long-run cost of providing these services, are discussed first. The household sample survey carried out to derive data on the Qs and to identify the factors influencing consumption is then described.

Identification of Public Services Through the National Budget

Public services can be thought of as a flow of services provided by a capital stock (or stock of infrastructure) and other variable resources such as labor and in-

15. The relation between the two definitions of the subsidy discussed earlier and the extra benefits to consumers out of public provision of the service are shown in figure 2-1.

Suppose D is the demand for the service and c and c' are the long-run marginal costs of providing the service by the public and by the private sector, respectively. Assume further that c' is greater than c (the public provision has a lower cost than the private substitute) and that the government prices the service at p, below its cost, c.

The implicit subsidy of having the service provided by the public sector with that pricing policy—instead of by the private sector—is equal to s' = s + Δc; s' is therefore equal to the pure pricing policy subsidy (the definition to be used in this study) plus the difference in cost between provision by the private and by the public sector.

The extra consumer surplus resulting from public provision of the service—when it replaces more expensive private sources of the service—is s' = [Q - Q2]. The one resulting from the pure pricing policy is equal to s [Q1 + 1/2 (Q2 - Q1)]. The concept of S used in this study is S = sQ2.
intermediate inputs. The long-run marginal cost of a unit of this flow includes the opportunity cost of capital required to provide this unit plus the associated variable cost.

To what extent does the government budget provide a measure for this long-run marginal cost? If the public sector hired or rented the capital stock or infrastructure from the private sector, the annual rental payments that would appear in the national budget could be used for this purpose. Since the public sector owns the stock, no rental payments are recorded in the budget. Only nonrental or running costs, that is, labor services and intermediate inputs, appear explicitly. The more labor intensive the production of a public service is, the more easily the long-run cost can be approximated by the cost data in the national budget.

The extreme, though not uncommon, example is that of a capital-intensive public service with negligible labor and intermediate input cost; for example, electricity derived from hydro sources. The national budget is unable to capture the true marginal cost of providing the service during a particular year. This cost can be estimated only from micro studies of an expansion of the given system.

The empirical implications of this are quite clear. By using the government's current expenditures as a substitute for the cost of the service, only the labor-intensive public services, that is, education and health, can be examined. Consequently, this study derives subsidy figures for only these sectors. The estimated figures will understate the true subsidy received by households to the extent that the rental cost of the capital stock in these sectors is not included in the budget.

No estimates are presented here for subsidies to capital-intensive public services. In the case of electricity, piped water, and sewerage, availability at one time and the supply-demand mechanism behind consumption are analyzed, and the income groups that benefited from expansions in the supply network over time are identified.

Some attempts were made to estimate the benefits of investment in roads, an important fraction of the national budget in Colombia. The mechanism by which this investment generates benefits, however, proved too complex to be measured by a household survey.

A different problem arises in dealing with the health services provided by the social security system for employees in the private sector. This system is financed largely by contributions from private employers and employees, with a minimal contribution from the central government. Although the measured contribution from the public sector is small, the system induces an important transfer across income groups.

Results

The survey was able to trace the beneficiaries of one-third of total government expenditure. The major expenditures accounting for this fraction are the public subsidies to the education and health sectors and the investment in electricity, water, and sewerage.

The total subsidy to education is distributed evenly across income quintiles; that is, the subsidy per household is constant across income groups. It results, however, from different subsidies to each educational level. The subsidy to primary education is progressive, whereas the subsidy to higher education is highly regressive; actually it is more unequally distributed than personal income. The health subsidy is also relatively similar across households, although it varies according to the source: the national health system has a progressive effect, whereas the social security system network favors the middle-income quintiles. The relative constancy of the subsidy per household across income groups does not hold when it is expressed in per capita terms. The reason for this is the difference in family size in each per capita income group. The poorest quintile of households accounts for 25.1 percent of the population, whereas the richest accounts for 15.4 percent, a difference of 60 percent in family size. The per capita subsidy to the richest quintile becomes 1.6 times larger than that to the poorest quintile.

For electricity, water, and sewerage, data were obtained on the distribution of households that had the service in 1974 and that received the service between 1970 and 1974. A comparison of the figures indicates the distributive direction of investment in those sectors over time.
The distribution of beneficiaries by quintiles in 1974 is quite similar across services. Between 25 and 30 percent of the households with services belong to the bottom 40 percent of households, whereas 50 to 55 percent belong to the upper 40 percent. Almost all consumers are concentrated in urban areas. The fact that low-income quintiles consume less of these services is more a result of those services being concentrated in urban areas—rural households being relatively poorer in the country distribution of income—than an intra-urban discrimination against low-income groups. The distribution of new beneficiaries is much more progressive. New investment in these sectors has been more redistributive than in the past. Part of this change results from the fact that investment had a lower "urban bias"; that is, a large fraction of the new beneficiaries have been rural households, particularly for electricity and piped water. Within the urban sector investment has also tended increasingly to benefit poorer households.

What factors explain the distribution of consumption of electricity, water, and sewerage services? A multivariate analysis was carried out to identify variables associated with the consumption of these services by urban households. A framework was developed whereby this association could be interpreted as a relation of causality. A situation in which the household does not use the service because of inaccessibility (the supply network is too far) is distinguished from a situation in which demand factors constrain the use of that supply. The Colombia data suggest that half the urban households without the service fall in the latter category. Therefore the study made a special effort to identify the extent to which variables such as per capita income, education, and migrant status affect the probability of a household's demanding the service.

The survey also identified the beneficiaries of other services and subsidies: street lighting and garbage collection services, educational fellowships, adult retraining courses at SENA (the Colombian Retraining Center), and the subsidy embodied in the loans from the Caja Agraria, the main public agency channeling subsidized credit to farmers. Although they are less important as a fraction of total government expenditure, the effect of these services and subsidies on particular groups can be important.

The survey was unable to identify the beneficiaries of investment in roads, a large item of public investment in Colombia. Questions regarding changes in the mode and time of travel did not yield significant results. Either investment in roads is basically beneficial as an intermediate input—rather than as a consumer good—or the questionnaire method used was not the proper technique to measure the benefit of this type of service.
The outstanding merit of an efficiently functioning price system is the attainment of an optimal allocation of resources. Once we define the properties of an optimal allocation, we can compare the actual situation with the optimal, and then proceed to the all-important task of devising remedial policies to reduce the divergence between the optimal and the actual.

For the economist, the optimal allocation of resources is defined in terms of a model of perfect competition. We now outline that model—first considering one perfectly competitive industry, and then examining an economy in which all industries are perfectly competitive.

A Perfectly Competitive Industry

The distinguishing characteristics of "perfect competition" are the following:

- Large numbers of small firms in the industry, with similarity in size and power.
- Freedom of entry into the industry. There are no natural obstacles to the entry of new firms into the industry, and new firms will enter if profits are above normal. ("Normal profits" cover the cost of capital and the opportunity cost of entrepreneurship; a sum above this would contain a producer's surplus.)
- Each seller is a "price taker," facing a price set by the market for its product. This is because there are many firms in the industry, each producing a standardized product, and no one supplying so large a proportion of the total as to affect the market price. So too is each buyer a "price taker," regarding the price as given and being able to buy as much as he likes at the market price.
- Information is perfect. There is no uncertainty about market conditions.

Given these characteristics and wishing to maximize its profits, a firm in a perfectly competitive industry must first determine how to produce each and every possible quantity of output at the lowest total unit cost (the least cost combination of factors), and then must select that particular quantity which would maximize profits. The details of this process are explained in Selections 3.1 and 3.2.

3.1 Efficiency of the Firm: Short Period

Suppose we are dealing with a competitive market for fish. How much of this commodity will be brought to market at each different level of market price? Firm A will bring so much to market at a particular price; Firm B will bring so much at this same price; Firm C will bring the amount shown on its supply curve; and so it goes. The total Q that will be brought to market at a given market P will be the sum of all the q's that firms will want to supply at that price. And similarly at any other price.

Summary: To get the aggregate SS supply curve for a good, we must add horizontally the SS supply curves of the independent producers of that good.

This is illustrated for two firms by figure 3.1. Recalling that the firms' momentary-run supply curves are defined as the inelastic supplies in a time period...
Figure 3-1. Market Supply Curve

(a) Firm A's Supply  
(b) Firm B's Supply  
(c) Their Aggregate Supply

Note: To get market supply curve, we add all firms' supply curves. At each price, such as $40, we horizontally add quantities supplied by each firm to get total market supply. This applies to any number of firms. If there were 1,000 identical firms, the market supply curve could be made to look just like the supply curve of each firm by a careful, thousandfold change of horizontal scale in the third diagram if no horizontal scale change is made, aggregate supply must look flatter than each firm's.

so short that no variability in output is possible, we can say:

To get the industry's vertical momentary curve $S_m$, add horizontally, at the same $P$, all firms' vertical momentary supply curves.

Now recall that in Selection 1.4 Marshall's "intermediate" or "short run" is defined as that period of time in which the firm is stuck with certain fixed commitments, but in which some variable factors of production can be altered so as to produce more output along the various firms' supply curves:

Again, to get the industry's short-run supply curve $S_s$, add horizontally, at the same $P$, the short-run supply curves of the fixed number of firms existing in that short run.

Our problem is to see how a firm's supply curve is determinable from its costs.

Definition of Marginal Cost Numerically and Graphically

Basic to industry supply is Marginal (or extra) Cost. Table 3-1 shows how we go about calculating Marginal Cost: By subtracting the $16,000 total cost of producing $q^* = 400$ units from the $16,040.05$ total cost of producing $q = 401$, we find $MC = 40.05$ for producing 1 more unit beyond $q^*$; to produce 1 less unit involves a difference, $MC = 39.95$, or $16,000 - 15,960.05$. So at $q^*$ itself we may, by disregarding or averaging over the trifling differences caused by the trifling lumpiness of units, estimate $MC = 40$.

Definition: Marginal Cost at any output level $q$ is the extra cost of producing one extra unit more (or less): it comes from subtracting total dollar costs of adjacent outputs.

Just as we can calculate $MC$ for $q^* = 400$, we can calculate it for any and every $q$. Table 3-1 had put a microscope on cost behavior around 400 and 401 units. To see the big picture, let us stand off and see how the Marginal Cost curve behaves at all levels of output.

$MC$ tends to be U-shaped: ultimately it is rising, even though there may be an initial phase in which $MC$ is falling.

Why can you expect Marginal Cost to be ultimately a rising curve? This takes us back to the law of diminishing returns. Behind the dollar costs of the firm lies the production relationship between the firm's output and the labor and other inputs it hires.

Suppose some factor is held fixed in the short run we are considering: it could be fixed land, or, in manufacturing, it could be fixed plant capacity. Sup-

1. WARNING: $MC$ is usually not the same as average cost per unit, which we get by dividing total cost by number of units produced: $MC$ is extra cost, or incremental cost, or differential cost; or, as we have seen from the use of the word "marginal" in connection with extra utility, the appropriate name is indeed Marginal Cost.

2. Figure 3-2 and the accompanying table measure $q$ in units of hundreds; hence, $q = 300, 399, 400, 401, 500$ in Table 3-1 would show in figure 3-2 as $q = 3, 3.99, 4, 4.01, 5$, and so on.
Marginal Cost. 1

Marginal Cost. 2

Marginal Cost. 3

Table 3-1. Calculation of Marginal Cost

<table>
<thead>
<tr>
<th>Quantity produced</th>
<th>Total cost</th>
<th>Marginal cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>(q)</td>
<td>(TC)</td>
<td>(MC)</td>
</tr>
<tr>
<td>399</td>
<td>$15,960.05</td>
<td>$39.95</td>
</tr>
<tr>
<td>400</td>
<td>16,000.00</td>
<td>40.00</td>
</tr>
<tr>
<td>401</td>
<td>16,040.05</td>
<td>40.05</td>
</tr>
</tbody>
</table>

Note: Marginal (or extra) cost can be shown numerically by subtraction of successive items. The difference in total dollar cost from producing an extra unit is found by subtracting adjacent items of total dollar cost in col. 2. At q = 400, MC = $40 to a high degree of approximation (as shown by the average of MC data).

pose that we get our varying amounts of q by hiring varying amounts of some input such as labor. If we can always buy labor at the same wage per unit, the only reason our marginal or extra cost of getting more q should rise would be because the extra product added by each successive unit of labor is going down. Hence, if we do get diminishing returns to the varying labor factor, we shall certainly get increasing Marginal Cost. Costs and productivity returns are merely opposite sides of the same relationship.

Why does MC decline at first, as shown in figure 3-2(b)? Recall that the law of diminishing returns tends ultimately to hold: at the beginning, it might be negated by a strong tendency toward increasing returns, owing to the economies of large-scale production associated with indivisibility of the process and chances to introduce more elaborate division of labor as scale expands. If at first we have strong increasing returns, we must at first have declining, rather than increasing, Marginal Cost.

3. Later we shall examine the behavior of Marginal Cost in the long run. Suppose we consider so extended a period of time that nothing can be regarded as fixed. Old plants can wear out and be replaced. New plants can be designed and built. Old land obligations can expire. New land contracts can be made. And so forth. In the long run, as a small firm, we may be able to buy all the factors of production in balance at unchanged prices. Now what will happen to long-run costs, particularly long-run MC, if the firm has no fixed factors and can enjoy "constant returns to scale"? (This is defined as a state where there is no reason for diminishing returns to operate, since all factors grow in balance, and where all economies of large-scale production have already been realized.) Answer: If long-run constant returns to scale hold, then doubling all inputs will exactly double their total dollar costs and will at the same time exactly double total output. Hence, there will be constant Marginal Cost, MC being horizontal rather than rising or falling.

We can summarize the relationship between the productivity laws of returns and the laws of Marginal Cost:

A tendency for varying factors to show diminishing returns when applied to fixed factors implies a tendency for MC to be rising. If at first there is increasing returns, there is at first declining MC—but ultimately diminishing returns and increasing MC.

How to Determine Maximum-Profit Competitive Supply by MC

It is evident that costs are vital determinants of how much a firm will be willing to supply. It would supply nothing if market P were too low to cover its out-of-pocket expenses, and would supply much if P were very high. To decide how much to supply at each market P, the firm will want to know the extra or marginal cost to it of each extra unit of q. Thus, consider Firm A in figure 3-1 which was shown supplying 4 (hundred) units at a market price of $40 per unit. Why does q = 4.00 and not 4.01 or 3.99?

At first you might be tempted to reply: "Firm A has no choice but q = 4 because that is all it can sell at $40—no more, no less." What is wrong with such an answer? It overlooks the fact that this is a model of perfect competition.

Definition: A perfect competitor is too small and unimportant to affect the market price. Like a wheat farmer, he is a "price taker" who can sell all he wishes at the ruling market price. In terms of demand elasticity, a perfect competitor faces a (virtually) horizontal dd demand curve for his product—his elasticity of demand is infinite.

Figure 3-3 shows the contrast between the industry demand curve DD, relating P to the sum of firm demands Q = q₁ + q₂ + ..., and the dd curve facing any one small competitor. If there are thousands of firms in the industry, the draftsman will have to rescale the horizontal axis and focus a microscope on point A of the industry demand curve, DD, in order to show dramatically how this sloped curve will reappear as the horizontal dd curve to the lilliputian eye of the firm.

Granted that a perfect competitor can sell any q he chooses at the going P, how does he pick his best q supply response? A perfect competitor picks the quantity he will supply by referring to his marginal-cost curve, so that P = MC. Why?

He will do this because he is interested in maximizing the total profit he can earn. Profit is the difference between the total revenue he receives from selling his output and the total cost incurred in producing that output. He increases his total profit so long as the extra revenue brought in from the last unit sold is greater than the extra
Figure 3-2. Marginal Cost

<table>
<thead>
<tr>
<th>Output (hundreds)</th>
<th>Marginal Cost</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>MC</td>
<td>TC</td>
</tr>
<tr>
<td>0</td>
<td>$55</td>
<td>$80</td>
</tr>
<tr>
<td>1</td>
<td>$30</td>
<td>$110</td>
</tr>
<tr>
<td>2</td>
<td>$25</td>
<td>$130</td>
</tr>
<tr>
<td>3</td>
<td>$20</td>
<td>$150</td>
</tr>
<tr>
<td>4</td>
<td>$30</td>
<td>$160</td>
</tr>
<tr>
<td>5</td>
<td>$50</td>
<td>$210</td>
</tr>
</tbody>
</table>

Note: Marginal Cost is the exact total cost. To find the MC of producing the fifth unit, subtract $160 from $210 to get $50.

In (a) a smooth curve has been drawn through the points of TC. In (b) a smooth MC curve has been drawn through the steps of extra cost. Ultimately MC is rising (because of diminishing returns), but it may at first fall, giving the curve a U-shaped contour.

cost which that last unit entailed. Total profit reaches its peak—when there is no longer any extra profit to be earned by selling extra output. The last little unit he produces and sells is just in balance as far as extra revenue and extra cost are concerned. What is that extra revenue? It is price per unit. What is that extra cost? It is marginal cost.

Specifically, in figure 3-1(a), why did you choose to produce quantity \( q^* = 4.00 \) (or 400 in the old units) at \( P = 40 \)? Only one answer is correct: Because the 401st unit would involve you in extra (or so-called marginal) cost of just over $40; and the 399th unit involved you in extra cost of just under $40; so best-profit \( q^* = 400 \) involves an extra or marginal cost just exactly equal to price \( P \) of $40.

Deriving the Firm’s Supply Curve from Its MC Curve

We have now demonstrated that a competitive firm is at its Maximum-profit position when it is definitely producing all units that had MC less than \( P \), and is definitely not producing further units for which MC is definitely greater than \( P \). Evidently, its Maximum-profit equilibrium comes when it follows the rule

\[
P = \text{Marginal Cost} \quad \text{or} \quad P = MC.
\]

This means that the firm’s supply curve is given by its rising MC curve as shown in figure 3-4. Thus, at the indicated horizontal \( d'd' \) level of $50, the firm will find its Maximum profit supply response at the intersection point A.

Alternatively, suppose the firm were faced by a horizontal \( dd \) at $40. Its Maximum-profit supply response is shown at the intersection point B. It happens, as our original cost table 3-1 shows, that the firm just breaks even there, covering all its long-run costs at this point, including both fixed costs and variable costs.

Or suppose the firm faced \( d''d'' \). At this price below $40, the firm cannot break even; but it does minimize its short-term losses (or maximize its “algebraic” profit) there at the C intersection on the MC curve. Thus, the firm’s rising MC curve does indeed constitute its competitive supply curve.
Total Cost and Short-Run Shutdown Conditions

Recall that earlier the "short run" was defined as that period of time in which certain equipment, resources, and commitments of the firms are fixed; but it is a period long enough for the firm to vary its output by hiring more or fewer variable factors of production, such as labor, raw materials, and so forth. It is certainly not a precise period of time that will be the same for all industries. Even within an industry, we can ask questions about "short-run" periods of different duration. At one ultrashort extreme, so many decisions have already been frozen as to make the resulting Marginal Cost curve practically a vertical and inelastic line. Or at the other extreme, we can permit so much time to pass as to let more and more of the equipment have a chance to wear away or be replaced, thereby making the resulting Marginal Cost curve almost as flat as it will be in the longest run, when no fixities are possible except those associated permanently with the management of the firm itself.

Now consider a firm making its short-run decisions. It has a certain "fixed cost": this is defined as the total of costs that will go on anyway because of its fixed commitments that are already frozen in the short run; examples would be bond interest, rentals, overhead salaries, and franchise taxes. The rest of its "total cost" is called "variable cost": this is defined as the sum of all costs that vary with output; examples are cost of materials and wages for workers on the production line.

But now consider the firm facing lower and lower \( P \). It has the option of producing nothing at all. How much will it then lose? With its revenue zero and all its fixed cost going on anyway, its shutdown loss will exactly equal its fixed cost. When \( P \) falls so low as to give it less revenue than the variable cost it incurs from producing positive \( q \), it will prefer to shut down completely: why should it produce if that means it incurs loss greater than the fixed cost it incurs when shut down? So this rule holds:

**Shutdown point.** At the critically low market price, where the firm just recovers its variable cost by producing, it will be on the verge of shutting down. Below that point, it will produce nothing at all.

Above that \( P \), it will produce along its short-run Marginal Cost curve. For, at such \( MC = P \) points, the firm will be getting something toward covering its fixed cost; and either it will be getting maximized positive profits, or, if \( P \) is below the Break-even point, at least the firm will be minimizing its losses (and, in that sense, maximizing its algebraic profit).

The location of the Shutdown point was shown in figure 3.4. The MC curve continues down below that point, but it no longer corresponds to an SS supply curve.

**Total Cost: Fixed and Variable**

Consider a typical firm that produces the output \( q \). At this stage, we do not care whether it is a perfect or imperfect competitor. At any one time, it has a certain state of technical knowledge, and it is confronted with the prices of the labor and other inputs it must buy. Now its accountants have been able to calculate what will be its total dollar costs for producing each different level of \( q \).

Table 3-2 shows the simplified Total Cost for each different level of output \( q \). Columns (1) and (4) are

4. These data depend upon engineering technology and upon the market prices of labor, land, fertilizer, and other factor inputs that the firm needs to produce each output of the good in question. Before its accountants and production men were able to write down the numbers in table 3-2, they had to make efficiency decisions in engineering, ensuring that the physical inputs could not be combined to give more physical output. The firm must also have made...
Table 3-2. Various Cost Concepts

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Fixed cost</th>
<th>Variable cost</th>
<th>Total cost</th>
<th>Marginal cost per unit</th>
<th>Average cost per unit</th>
<th>Average fixed cost per unit</th>
<th>Average variable cost per unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>FC</td>
<td>VC</td>
<td>TC</td>
<td>MC</td>
<td>AC = \frac{TC}{q}</td>
<td>AFC = \frac{FC}{q}</td>
<td>AVC = \frac{VC}{q}</td>
</tr>
<tr>
<td>0</td>
<td>55</td>
<td>0</td>
<td>55</td>
<td>34</td>
<td>Infinity</td>
<td>Infinity</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>30</td>
<td>85</td>
<td>27</td>
<td>85</td>
<td>55</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>55</td>
<td>110</td>
<td>22</td>
<td>55</td>
<td>27.5</td>
<td>27.5</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>75</td>
<td>130</td>
<td>21</td>
<td>43%</td>
<td>18%</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>105</td>
<td>160</td>
<td>40</td>
<td>40%</td>
<td>13%</td>
<td>26%</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>155</td>
<td>210</td>
<td>60</td>
<td>42</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>225</td>
<td>280</td>
<td>80</td>
<td>46%</td>
<td>9%</td>
<td>37%</td>
</tr>
<tr>
<td>7</td>
<td>55</td>
<td>-</td>
<td>370</td>
<td>100</td>
<td>52%</td>
<td>7%</td>
<td>45</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>-</td>
<td>480</td>
<td>120</td>
<td>60</td>
<td>6%</td>
<td>53%</td>
</tr>
<tr>
<td>9</td>
<td>55</td>
<td>555</td>
<td>610</td>
<td>140</td>
<td>67%</td>
<td>6%</td>
<td>61%</td>
</tr>
<tr>
<td>10</td>
<td>55</td>
<td>705</td>
<td>760</td>
<td>150</td>
<td>76</td>
<td>5%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Note: From schedule of firm's total cost, all other costs can be computed. All the costs can be calculated from col. (4)'s rising TC.Cols. (5) and (6) are the important ones to concentrate on: incremental marginal cost is calculated by subtraction of adjacent rows of TC; the numbers of smoothed MC come from figure 3-5(b). In col. (6) note the point of minimum cost of $40 on the U-shaped AC curve. (Realize why the MC = AC at the minimum.)

a. Minimum level of average cost.

the crucial ones, showing that TC goes up as q goes up. This is natural because it takes more labor and factor inputs to produce more of a good, and these extra factors involve an extra money cost. It costs $110 in all to produce two units, $130 to produce three units, and so forth.

Columns (2) and (3) break down Total Cost into two components: total Fixed Cost, FC; and total Variable Cost, VC. Figure 3-5(a) shows the breakdown.

Even when the firm produces zero output, it must honor its short-run commitments (contractual rentals, watchmen's pay) and continue to incur its total Fixed Cost of $55. By definition, FC is the amount of cost that goes on independently of output; so it remains constant at $55 in Column (2). Another name for Fixed Cost is overhead cost.

Column (3) shows total Variable Cost. By definition, VC begins at zero when q is zero. It is the part of TC that grows with output; indeed, the jump in TC between any two outputs is the same as the jump in VC. Why? Because FC stays constant at $55 throughout and cancels out in any such comparison. (Fill in by subtraction the missing VC data of the third column.)

Definitions: "Total Cost" represents lowest aggregate dollar expense needed to produce each level of output q. TC rises as q rises.

"Fixed Cost" represents the total dollar expense that goes on even when a zero output is produced. It is often called "overhead cost" and usually includes contractual commitments for rental, maintenance, depreciation, overhead salaries and wages, etc. It is a sunk cost that is quite unaffected by any variation in q; in the time period for which it is sunk, the only rule is this: Disregard Fixed Cost because FC cancels completely out of every decision.

"Variable Cost" represents all items of TC except for FC—as, for example, raw materials, wages, and fuel. Always, by definition,

\[ TC = FC + VC. \]
Figure 3-5. Total Cost Curve

(a) Total, Fixed, and Variable Costs

(b) Average Cost

Note: Total Cost curve gives all other curves.

(a) The total Fixed Cost curve is horizontal by definition. Adding on top of it the rising total Variable Cost gives the rising Total Cost curve.

(b) The curve of Marginal Cost falls and ultimately rises. MC is shown as a smooth curve after the steps of incremental cost are smoothed out; the numbers from the smooth MC curve are given in col. (5) of table 3-2 (and also correspond to the slope of TC curve shown above).

By dividing TC by \( q \), we can plot Average Cost: \( AC = \frac{TC}{q} \). Similarly, AFC comes from \( FC/q \), and AVC comes from \( VC/q \). At any point we can add these two curves and also get the AC curve at that point.

Note that MC intersects the U-shaped AC at AC's minimum. This is no coincidence. To the left of M, MC < AC, and hence is pulling AC down. To the right, MC > AC, and hence is pulling AC up. At the M minimum point, MC = AC; hence AC is horizontal there, being neither raised nor lowered by the equivalent MC. Also, MC cuts the AVC curve at its bottom.

Note that TC and VC always show exactly the same increments of MC as \( q \) changes, because FC is a strict constant.

Review carefully how in figure 3-5(a) the rising TC curve is shown broken down into its constant FC and rising VC components.

Marginal Cost Review

We saw how Marginal Cost is defined as the increment of Total Cost that comes from producing an increment of one unit of \( q \). (Recall that “marginal,” whether applied to utility, cost, or anything else, always means “extra” in economics.)

The MC numbers in Column (5) of table 3-2 come from subtracting the adjacent TC numbers in Column (4). Thus MC is $30 in going from 0 to 1 unit of \( q \) (i.e., $85 - $55 = $30). MC is seen to be $110 - $85 = $25 in going from \( q = 1 \) to \( q = 2 \). MC is $20 for the third unit of \( q \), $30 for the fourth, and thereafter rising steadily until it is shown as $150 in going from \( q = 9 \) to \( q = 10 \). (What is MC in going from \( q = 5 \) to \( q = 6 \)?)

Instead of getting MC from the TC column, we could as easily get the MC number by subtracting each VC number of Column (4) from the row below it. Why? Because Variable Cost always grows exactly like Total Cost, the only difference being that it must—by definition—start out from 0 rather than from the constant FC level. (Check that $30 - 0 = 85 - 55$, and $55 - 30 = 110 - 85$, ...)

Figures 3-5(a) and (b) show the behavior of Total Cost and Marginal Cost. Note that the \( q \) axes of the graphs are just lined up so that the eye can see the correspondence between TC and MC.

This example shows MC to be U-shaped—at first falling, but ultimately rising. At first there may be great economies in using some or all of the productive factors on a larger scale; and so MC at first falls down to a minimum positive number before again rising. If we stick to the short run where some factors of production are fixed, ultimately the old law of diminishing returns will operate to reduce the extra product that comes from adding equal physical and dollar increments of the varying factors onto the fixed factors. So the cost of getting extra product will ultimately become more expensive, and the short-run curve of Marginal Cost will ultimately be rising.

Marginal Cost has many uses. We saw that the rising MC curves of the firms provide us with the rising short-run SS supply curves which we sum horizontally to get the industry's short-run SS supply curve. We shall see that any firm—monopolistic or competitive—will find its Maximum-profit equilib-
rium by nicely balancing its extra cost against its extra revenue (i.e., by finding an intersection of its Marginal Cost curve and what will be defined as its "Marginal Revenue" curve).

**Average or Unit Cost**

But first turn back to Column (6) of table 3-2. This gives Average Cost (per unit), which is simply the Total Cost divided by the number of \( q \) units produced.

\[
\text{Average Cost} = \frac{\text{Total Cost}}{\text{output}} = \frac{TC}{q} = AC.
\]

In Column (6), when only 1 unit is produced, Average Cost has to be the same as Total Cost, or \( 85/1 = 85 \). But for \( q = 2 \), \( AC = TC/2 = 110/2 = 55 \), as shown. Note that Average Cost is, at first, falling lower and lower. (We shall see why in a moment.) But AC reaches a minimum of \( 40 \) at \( q = 4 \), and then slowly rises.

Figure 3-5(b) gives a careful plotting of U-shaped AC, nicely arranged below the TC it came from. We can now break down Average Cost into its two components, fixed and variable, just as earlier we had the breakdown of TC into FC and VC. By dividing each of the last two by \( q \), we get Average Fixed Cost, \( AFC = FC/q \) of Column (7); and Average Variable Cost, \( AVC = VC/q \) of Column (8).

**AVERAGE FIXED COST**

Since total Fixed Cost is a constant, dividing it by \( q \) gives in Column (7) a steadily falling Average Fixed Cost curve. The dashed AFC curve in figure 3·5(b) looks like a unitary demand curve, or hyperbola, that approaches both axes: it drops lower and lower, approaching the horizontal axis as the constant FC gets spread over more and more units. If we allow fractional and zero units of \( q \), AC starts infinitely high, as the finite FC is spread over tinier and tinier units of \( q \).

**AVERAGE VARIABLE COST**

AVC of Column (8) and figure 3-5(b) at first falls and then ultimately rises. We could have predicted this U-shaped behavior of AVC from the U-shaped behavior of MC. When MC at first falls, each new \( q \) is pulling down the Average Variable Cost calculated over all the items.

**THE POINTS OF MINIMUM AVERAGE COST**

Figure 3-5(b) is an important economic diagram. Fix it on your eye's retina. Note particularly the typical U shape of the AC curve. The AC curve is always pierced at its minimum point by the rising MC curve.

This is no coincidence. And now we can explain why this has to be the case. Any average curve is pulled downward when MC is less than AC: If the last increment of cost is less than the average of all previous ones, it must pull the average down! But when MC gets as big as AC, AC no longer is pulled down; it now turns sideward or level. Then, if MC rises above the AC, it must of course pull AC up. So at the point where rising MC = AC, and only at that point, shall we find the point of Minimum AC.

Summary: So long as Marginal Cost is below Average Cost, it is pulling Average Cost down; when MC gets to be just equal to AC, AC is neither rising nor falling and is at Minimum AC; after MC is above AC, it is pulling AC up. Hence:

At bottom of U-shaped AC,

\[
MC = AC = \text{Minimum AC}.
\]

Likewise, the MC curve cuts the AVC curve at the bottom of its U, pulling it down before this point because MC < AVC and pulling it up beyond that point because MC > AVC.

Now that we know how the MC curve intersects the AC and AVC curves where their U's bottom out, we can describe exactly how the firm's Shutdown and Break-even points of figure 3-4 had been determined. The Break-even point of long-run no-profit competitive equilibrium is seen here in figure 3-6 to
Figure 3-7. Finding Maximum-profit Equilibrium

(a) Long-run Cost Curve

$\text{Costs per unit}$

\[ LAC \]

\[ SAC' \]

\[ SAC'' \]

\[ SAC''' \]

\[ 0 \]

\[ q \]

\[ \text{Quantity} \]

Note: In the long run, a firm can choose its best plant sizes and its lower-envelope curve. (a) LAC is the "envelope" or lower frontier of the three possible choices of plant. (b) There is now an indefinite number of choices, and we get LAC as a smooth envelope. In the usual way we derive from LAC curve its marginal curve, LMC.

be at the bottom of the U-shaped AC curve, in accordance with

\[ P = MC = \text{Minimum AC}, \text{ in long-run equilibrium of zero excess-profits}. \]

Likewise, the Shutdown point would come at M', once the horizontal dd of the firm had fallen to a level of P so low as to just cover minimum AVC. A price below this level would cause the firm to produce zero output.

Long-Run Planning Envelope Curve

We now have all the technical apparatus of the various cost concepts needed to permit us to tackle the problem of how any firm will find its Maximum-profit equilibrium. But one last technicality is needed to explain how a firm may in the longest run be able to have lowest costs through adapting and varying the size of its plant.

Recall that, once the firm has a fixed plant, it has a short-run U-shaped AC curve (call it SAC to emphasize its short-run nature). If the firm builds a larger plant, the new SAC curve must be drawn farther to the right. Now, suppose the firm is still in the planning stage, still quite uncommitted by any fixed obligations. It can write down all possible different U-shaped SAC curves, and then choose to select for each prescribed output the SAC that gives it the lowest costs. As q changes permanently, the firm hops to a new SAC curve.

Figure 3-7(a) shows how, in the longest run, the firm selects SAC' for low q; for intermediate q, it does better to plan to use SAC''; for still larger q, SAC''' leads to lowest costs. The heavy curve of Long-run Average Cost (LAC) is composed of the three lowest branches. Figure 3-7(b) shows the same lower limit in the case where the firm has choice of infinitely many smooth short-run AC curves (SAC, not AVC): now LAC is the U-shaped smooth "lower-envelope" curve; and its well-behaved LMC provides the firm's long-run Marginal Cost curve, emerging from the LAC minimum point with a gentler slope than the short-run SMC curve there.

Efficiency of the Firm: Long Run

Over the short run, firms cannot escape from their plant capacity or fixed costs; they therefore maximize their profits by equating Marginal Revenue (MR) to Short-Run Marginal Cost (SRMC). Because demand is perfectly elastic for each firm (firms are price takers that cannot affect market price by changing their supply), the price is also equal to Marginal Revenue. Thus, \( P = MR = SRMC \) (see figure 3-4). But if the SRMC is above the the firm's Average Total Cost (ATC) there will be abnormal profits (normal profit is already included as a cost to capital and
entrepreneurship in the ATC). New firms will then enter the industry. As they do so, the demand for each firm will decrease, and the average cost will rise as firms bid for the scarce resources and prices of inputs rise. As a result, the abnormal profits are whittled away as supply increases. Over the long run—after short-run fixed costs have been amortized, plants have expanded, and new firms have entered the market—Price, Marginal Revenue, Long-Run Marginal Cost, and Minimum Long-Run Average Total Cost will be equal for each firm, that is, \( P = MR = LRMC = \text{minimum LATC} \) (figure 3-6.)

These are equilibrium conditions, for if they hold, no new firms will enter and old firms will no longer expand output. Each firm will be producing its best output in the sense of maximizing normal profits, but there will be no abnormal profits. The higher-cost, inefficient firms will be driven out of the industry and the invisible hand will "strangle" those firms that cannot reduce their ATC and sell at the market price. The lower-cost, more efficient firms will have to pay a rent to their unique resources to retain them in the firm when other firms bid for these unique resources. The inclusion of rent to these resources will raise the ATC curve of the lower-cost firms up to the same level as other firms. While each firm covers its normal profit, as given by the condition \( P = \text{minimum LATC} \), there are no abnormal profits. By producing at its minimum Long-Run Average Total Cost, each firm is minimizing costs of production in terms of productive effort.

Finally, the equality between Price and Long-Run Marginal Cost indicates that the marginal valuation of the buyer, as represented by the price, just equals the Marginal Cost to the economy of producing this output, as represented by the Long-Run Marginal Cost. (The assumption, on which the foregoing is based, that all costs are reckoned in the market, and that social costs do not differ from private costs, is discussed later, in detail, in this chapter.) By allocating resources in an optimal fashion, output is ideal: through the forces of competition, output is being produced at the lowest possible cost, and the amount of output is neither too large nor too small. No other allocation would leave consumers and producers as well off.

A Perfectly Competitive Economy

In an economy in which all industries are perfectly competitive, in each industry the long-run equilibrium conditions are \( P = MR = MC = \text{minimum LATC} \). Under these conditions, the allocation of resources in the economy is optimal and economic efficiency is attained.

In Selection 1.12, equilibrium was expressed as the Marginal Rate of Transformation (MRT) in production between any two industries being equal to the consumers' Marginal Rate of Substitution (MRS) between the two industries. The price ratio between the two industries indicates the rate at which producers can transform one product into another, and this rate is equivalent to the rate at which consumers prefer to substitute one product for another. The economy is, so to speak, on its production possibility frontier, and each consumer is as high as possible on his scale of preferences.

These conditions \( P = MR = MC = \text{minimum LATC} \) are also referred to as Pareto optimality, indicating that after the optimal resource allocation is achieved it would be impossible to make anyone better off without making someone else worse off. In a perfectly competitive economy, voluntary exchange would achieve
Pareto optimality: as long as the parties to transactions will benefit, they will voluntarily enter into such market transactions.

As explained in Chapter 1, any point on the production possibility frontier represents a position of technical efficiency—a maximum possible combination of goods being produced with given resources. But the particular point reached on the production frontier under conditions of perfect competition represents a position of economic efficiency. Production is not only maximized, it is also optimized. Producers maximize the value of their output by producing the combination of products that achieves an equality between Marginal Rate of Transformation between any two products and the ratio of their market prices. The production-optimizing position depends on exchange relations in the market, not merely technological relations in production. Not only is a maximum possible combination of goods being produced, but also the distribution of goods among consumers is such that no further voluntary exchange would occur. Each consumer is equating the Marginal Rate of Substitution between any pair of commodities he consumes to the ratio of their market prices. Therefore, the Marginal Rate of Transformation equals the Price ratio equals the Marginal Rate of Substitution. The exchange relations then constitute an optimal allocation of resources or economic efficiency or, for short, Pareto optimality.

In such a situation, the following conditions obtain: (1) the marginal rates of substitution between any pair of factors are the same in all industries in which they are used; (2) the price of a factor equals the value of its marginal product, and this is equal in all industries; (3) the Marginal Rates of Transformation in production of different commodities are equal to the rates at which they exchange on markets and to their Marginal Rates of Substitution in consumption; and (4) the Marginal Rates of Substitution of any pair of commodities are the same for all individuals consuming both products.

These conditions of economic efficiency denote a resource allocation that cannot be improved upon in the sense of increasing the real income of somebody without reducing the real income of someone else.

International Economic Efficiency

The same principles that apply within a nation for optimal resource allocation also apply to the entire world economy and resource allocation among nations. Internationally, free trade markets are equivalent to competitive markets. If firms were perfectly competitive, and there were no impediment to world trade, then the resultant resource allocation would again be optimal. Free trade prices (in the absence of market failure as discussed later) would be efficiency prices.

This conclusion can be derived from the principle of comparative advantage. Although the modern theory of comparative advantage has become extremely refined to incorporate any number of countries, commodities, and factors of production, its basic logic is easily understood from the original presentation by David Ricardo (1817). Adam Smith had earlier argued for free trade on the basis of absolute advantage—the capacity of a country to produce its export commodities at absolutely lower cost, in terms of real resources used, than they could be produced abroad and to import those commodities that another country can produce at absolutely lower real cost.
In Smith’s words,

It is the maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy. The tailor does not attempt to make his own shoes, but buys them of the shoemaker. The shoemaker does not attempt to make his own clothes, but employs a tailor. The farmer attempts to make neither the one nor the other, but employs those different artificers. All of them find it for their interest to employ their whole industry in a way in which they have some advantage over their neighbours, and to purchase with a part of its produce, or what is the same thing, with the price of a part of it, whatever else they have occasion for.

What is prudence in the conduct of every private family, can scarce be folly in that of a great kingdom. If a foreign country can supply us with a commodity cheaper than we ourselves can make it, better buy it of them with some part of the produce of our own industry, employed in a way in which we have some advantage. The general industry of the country, being always in proportion to the capital which employs it, will not thereby be diminished . . . ; but only left to find out the way in which it can be employed with the greatest advantage. It is certainly not employed to the greatest advantage, when it is thus directed towards an object which it can buy cheaper than it can make . . .

The natural advantages which one country has over another in producing particular commodities are sometimes so great, that it is acknowledged by all the world to be in vain to struggle with them. By means of glasses, hotbeds and hotwalls, very good grapes can be raised in Scotland, and very good wine too can be made of them at about thirty times the expense for which at least equally good can be brought from foreign countries. Would it be reasonable law to prohibit the importation of all foreign wines, merely to encourage the making of claret and burgundy in Scotland? But if there would be a manifest absurdity in turning towards any employment, thirty times more of the capital and industry of the country, than would be necessary to purchase from foreign countries an equal quantity of the commodities wanted, there must be an absurdity, though not altogether so glaring, yet exactly of the same kind, in turning towards any such employment a thirtieth, or even a three hundredth part more of either. Whether the advantages which one country has over another, be natural or acquired, is in this respect of no consequence. As long as the one country has these advantages, and the other wants them, it will always be more advantageous for the latter, rather to buy of the former than to make. It is an acquired advantage only, which one artificer has over his neighbour, who exercises another trade; and yet they both find it more advantageous to buy of one another, than to make what does not belong to their particular trades.

Thus, just as the division of labor increases output in domestic trade, so too would the international division of labor, based on absolute advantage, augment the "value of annual produce." To Smith, international trade was simply a specific manifestation of the general principle of exchange, and international specialization followed the dominant principle of the division of labor. Promoting this specialization were differences in climate, qualities of soil, and other natural or acquired advantages, which resulted in absolute differences in real costs of production. As the cost of growing grapes in Scotland illustrates, a country does not import because it is physically impossible to produce the importable commodity, but rather because the real cost of doing so would exceed what it would cost to specialize on its low-cost export commodity and exchange it for the import.

Comparative Advantage

But what if a country is able to produce every commodity at an absolutely lower real cost than another country (instead of just one or two commodities, as in Smith's example)? Would there then still be a basis for trade? The answer is "yes"—according to David Ricardo's celebrated principle of comparative advantage. Under conditions of free trade, a country will specialize in the production and export of those commodities it can produce with greatest comparative advantage, i.e., those commodities for which its costs are comparatively lowest, and will import commodities in which it has a comparative disadvantage, i.e., those commodities it can produce only at high relative cost.

To illustrate the meaning of comparative costs, Ricardo focused on a simple example of two countries, two commodities, and one factor of production, labor—a 2 × 2 × 1 model. Consider this example below. In Portugal (P), 1 unit of wine (1w) costs 80 units of labor and 1 unit of cloth (1c) costs 90 units of labor, while in England (E), 1w costs 120 units of labor and 1c costs 100 units of labor.

<table>
<thead>
<tr>
<th>Country</th>
<th>Units of Labor Required for Producing 1 Unit of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal</td>
<td>wine: 80, cloth: 90</td>
</tr>
<tr>
<td>England</td>
<td>wine: 120, cloth: 100</td>
</tr>
</tbody>
</table>

In contrast, Smith's case of absolute advantage would be represented by figures such as the following. This table represents "absolute differences in cost," with P having an absolute advantage in w, and E having an absolute advantage in c.

<table>
<thead>
<tr>
<th>Country</th>
<th>Units of Labor Required for Producing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal</td>
<td>1w: 100, 1c: 200</td>
</tr>
<tr>
<td>England</td>
<td>1w: 120, 1c: 100</td>
</tr>
</tbody>
</table>

Another possible case is shown below. In this special case, P has an absolute advantage in both commodities, but by the same amount: it is twice as productive in both w and c. This is a condition of "equal differences in cost," and there would be no basis for foreign trade. For in each country, 1.2c could be produced by sacrificing the production of 1w, and there could be no international exchange ratio established between the domestic cost ratios such that benefit would ensue to each country from trading. (This will become clear after the next case of "comparative differences in cost.")

<table>
<thead>
<tr>
<th>Country</th>
<th>Units of Labor Required for Producing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal</td>
<td>1w: 60, 1c: 50</td>
</tr>
<tr>
<td>England</td>
<td>1w: 120, 1c: 100</td>
</tr>
</tbody>
</table>

It is, however, scarcely conceivable that a country would be equally inefficient (or equally efficient) in all activities. Instead it is likely that there will be a difference between the ratios of the costs in the different countries. It will then be cheaper to acquire goods through trade than to produce them at home when the terms of international exchange are more favorable than the ratio of costs at which they could be produced at home.

This returns us to Ricardo's example: the interesting question is whether P would export one commodity and import another commodity even though it
could produce both commodities at absolutely lower costs than E. Ricardo answered that

It would be . . . advantageous for her [Portugal] to export wine in exchange for cloth. This exchange might even take place, notwithstanding that the commodity imported by Portugal could be produced there with less labour than in England. Though she could make the cloth with the labour of 90 men, she would import it from a country where it required the labour of 100 men to produce it, because it would be advantageous to her rather to employ her capital in the production of wine, for which she would obtain more cloth from England, than she could produce by diverting a portion of her capital from the cultivation of vines to the manufacture of cloth. [Italics added.]

In drawing this conclusion, Ricardo assumes domestic mobility of resources, but international immobility (so that labor does not migrate to the higher productivity country), production subject to constant returns to scale (constant costs in each activity), and perfectly competitive markets. To emphasize the “real” theory of international trade equilibrium, Ricardo abstracts from any monetary imbalance and assumes that exports must equal imports, with no capital flows.

The Ricardian example shows that P has an absolute advantage in both w and c, but a comparatively greater advantage in w, since 80/120 < 90/100; P is superior to E in the production of both commodities, but “more superior” in w than in c. (It does not matter whether the cost ratios we compare are 80/120: 90/100 or 80/90: 120/100.) In both commodities E has an absolute disadvantage, but a lesser disadvantage in c. If labor is perfectly mobile within each country and prices equal labor costs, then in the pre-trade situation, the domestic exchange ratio between the two commodities would be 1w: 0.88c in P, and 1w: 1.2c in E. (The exchange ratio in physical units is the reciprocal of the price ratio.) Thus, w is comparatively cheap in P, whereas c is comparatively cheap in E.

This can be represented diagrammatically in figure 3-8, in which the production-possibility frontier for E is denoted by MD. The location of M and D are determined by the total supply of labor in E. The points along MD indicate maximum possible combinations of w and c that could be produced in E with the existing labor supply. The slope of MD shows the rate at which the resources embodied in the production of one commodity can be transformed into the production of another commodity. In this case, the labor that produces 1.2c can be transformed into the production of 1w by sacrificing the production of 1.2c. The condition of constant costs in each activity makes MD a straight line. The slope of the production frontier is termed the “marginal rate of domestic transformation” (MRT_d) because it shows the rate at which one commodity can be “transformed” into another in the production process.

We can also view the slope, or MRT_d, as identifying the marginal cost of one commodity in terms of the other or as denoting the opportunity cost of one commodity in terms of the other. If product markets are perfectly competitive, prices must equal marginal costs. Therefore, the slope of the production possibility curve also indicates the ratio of marginal costs of the two commodities. The marginal cost, or opportunity cost of producing one more unit of w, is the forgone production of 1.2 units of c.

Figure 3-8. Pre-trade and Post-trade Positions for England

The slope of E's production-possibility frontier represents its comparative cost ratio; P will have a different production-possibility frontier, determined by its total supply of labor and ratio of marginal costs, such as that in figure 3-9. Given different production frontiers in the two countries, a basis for international trade arises from these relative productivity differentials or differences in the countries' comparative cost ratios. After opening E and P to foreign trade, some exchange ratio on world markets will be established. This international exchange ratio must lie within the range of the domestic exchange ratios, between Portugal's 1w: 0.88c ratio and England's 1w: 1.2c (ignoring transportation costs). Otherwise one country would not trade. If, for example, the international exchange ratio were 1w: 1.5c, E would not trade c for w at this international rate when it could produce 1w by forgoing the production of only 1.2c at home. Even though P would have a high demand for c at this rate, a supply of c would not be forthcoming from E until the rate became more favorable than 1w: 1.2c for E. Exactly where the rate will be established between the domestic exchange ratios will depend on world demand and supply conditions.

For a balanced trade, the terms must be such that the amount of w that E would purchase with c, at those terms, would cost the same amount of c that P would want to buy with w. The exact position of these equilibrium terms of international exchange will depend on the intensity and elasticity of E's demand for w and P's demand for c.

Of all the possible international exchange ratios between P's and E's comparative cost ratios, we have taken MF in figure 3-8 and D'F in figure 3-9 as the equilibrium terms (MF and D'F have the same slope). At these terms of trade, the value of exports equals the value of imports for E and P.
Let us say that demand and supply conditions determine the equilibrium international ratio as $1w: 1c$. Then $E$ would specialize in the production of $c$, its comparative advantage commodity (lesser disadvantage), and export $1c$ for $1w$, thereby gaining 0.17 units of wine for each unit of cloth exported (the pre-trade exchange ratio is $1c: 0.83w$). Alternatively, $E$ would acquire $1w$ at a lower real cost than in the pre-trade situation (a saving of 0.2 units of $c$ because the pre-trade exchange ratio is $1w: 1.2c$). And $P$ would specialize in $w$, its comparative advantage commodity (greater advantage), thereby gaining 0.12 units of cloth for each unit of wine exported (the pre-trade exchange ratio is $1w: 0.88c$). Alternatively, $P$ would acquire the same quantity of $w$ and $c$ as produced before trade at a lower total real cost after trade.

To summarize: The Ricardian explanation of the pattern of international specialization rests on the principle of comparative advantage. Differences between countries in the comparative productivity of labor (technological differences in the production functions) give rise to comparative differences in real costs and to differences in the countries’ pre-trade commodity price ratios. When the countries are opened to foreign trade, a common international price ratio can then be established between the domestic price ratios. At home, the cost of one commodity in terms of another is greater than on world markets, and the country therefore specializes in producing the commodity that uses fewer resources than would be needed to produce domestically the commodity it imports. It exports the commodity in which it has a comparative advantage and imports the commodity in which it has a comparative disadvantage. Thus, the cost of “indirectly
producing" imports through specialization on exports is less than if the country produced the imports directly at home.

Exports can be viewed as the intermediate goods used for the "production" of imports. Foreign trade is like an industry that uses exports as inputs to produce imports as output. Indeed, in a centrally planned economy, in which the foreign trade sector is treated as an industry, the principle of comparative advantage should be adopted as the efficiency rule for determining what to export and what to import. In following its comparative advantage, each country maximizes output (imports) per unit of input (exports).

If the number of commodities and number of countries are increased, the model becomes more complicated, but the logic of the analysis remains the same. Commodities will be ranked by their comparative factor-productivity ratios such that each of a country's exports will have a higher factor-productivity ratio than each of its imports. Any number of commodities in any number of countries can be arranged in a chain of declining comparative advantage. In this chain, the position of the dividing line between exports and imports will depend on demand conditions for each country's products and on the equilibrium conditions that world demand equals world supply and that value of exports equals value of imports for each country.

Although the Ricardian analysis establishes a basis for international specialization according to comparative differences in real cost, it is, of course, necessary to translate the comparative differences in real cost into absolute differences in money prices. Obviously, any buyer of imports will purchase imports only if the absolute money price is less than the price of a domestic substitute good. This translation into absolute differences in money price will follow, provided the relative wage differentials are within the relative productivity differentials. If, for example, labor is three times more productive in country I than in country II, the absolute money price of the exports from I will be less than the absolute money price of the substitute commodity in II, provided the money wage rate is not more than three times higher in I than in II. This can be illustrated by comparing the results in table 3-3(a) and (b).

Under competitive conditions, the wage differences must be within the productivity differences (real cost), or else only one country would be able to export, and its wage rates would be bid up to compensate for this. Or the exchange rate would have to be altered to preserve the condition of balanced trade.

We can more readily understand how comparative differences in real cost become translated into differences in absolute money prices if we appreciate the relationships between productivity, wage rates, foreign currency exchange rates, and the resultant changes in absolute prices. Suppose country I is more productive than country II in every commodity, and that I can therefore readily export to II, but II cannot export to I. But when the demand increases for I's exports, its wage rates will also increase, thereby raising its absolute prices and reducing its competitive advantage. The fall in demand in II may also lower wage rates and absolute prices in II, thereby allowing II to acquire a competitive advantage in another commodity. Or as the demand rises for I's currency on foreign exchange.


markets, the price of I's currency in terms of II's currency will rise (I's currency will appreciate), and this will cause the demand for I's exports to decline and will raise I's demand for imports. Thus, through wage changes or foreign exchange rate movements, the goods of country I will rise in price and those of country II will fall.

Similarly, even if one industry in I is so competitive that it dominates the world export market (say, Japanese electronics), the resulting exchange rate appreciation (say, a change from ¥ 200: $1 to ¥ 150: $1) will make other export industries in I less competitive. Because of variations in wage rates and exchange rates the underlying differences in comparative real costs are reflected in differences in absolute prices.

Aside from explaining the pattern of international specialization, the Ricardian model—despite its highly simplified statement—allows us to draw a number of policy conclusions. We can, for instance, immediately think of several policy decisions that could increase a country's exports. Recognizing that another country will import if the money price of imports is absolutely lower than the price of domestic substitutes, we can summarize the foregoing analysis in the relationship below.

Exports of the first country could be increased if the second country lowered its tariffs (the free-trade assumption), transportation costs fell (the no transportation cost assumption), or the money cost of producing the first country's exports declined because of an increase in factor productivity or a lower factor price (the basis for translating comparative differences in labor costs into absolute differences in money costs and money prices). Alternatively, the country could allow its currency to depreciate on the foreign exchange markets so that the prices of its exports would be lowered in terms of foreign currency.

---

Table 3-3. Real Costs and Absolute Money Prices

<table>
<thead>
<tr>
<th>Country</th>
<th>Amount of labor required for producing 1 unit of foreign currency exchange rate</th>
<th>Wage rate per labor unit at existing foreign currency exchange rate</th>
<th>Money Price</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Country X</td>
<td>Y</td>
<td>X</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>10</td>
<td>$2</td>
</tr>
<tr>
<td>II</td>
<td>6</td>
<td>30</td>
<td>$1</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>10</td>
<td>$3</td>
</tr>
<tr>
<td>II</td>
<td>6</td>
<td>30</td>
<td>$1</td>
</tr>
</tbody>
</table>
In explaining the basis of trade, the theory of comparative costs has a remarkably high degree of generality. It can be applied widely to problems of resource allocation—whether for an individual, for a country or for several countries or within a multi-product enterprise. Its logic is institution-free and applicable to any decision mechanism that must exercise the logic of choice.

Although Ricardo's logic is powerful, his technical analysis leaves much to be desired by modern standards. Economists have therefore elaborated upon the Ricardian analysis—introducing many factors of production, various cost conditions, and focusing on differences in relative factor endowments, technology, and demand conditions as the determinants of comparative cost ratios. All these elaborations, however, only reinforce the basic principle that can be derived from Ricardo—namely, a regime of free trade would be based on the pattern of comparative advantage, and this would result in the optimal allocation of resources internationally.

Optimality of Free Trade

The optimality of free trade can be demonstrated in the following way. First, consider a change from no trade to free trade. In figure 3-10, let \( P_h \) represent the domestic price before trade, and assume that there is a prohibitive tariff—say, at the specific rate of \( P_h P \), per unit of import, or the \( \textit{ad valorem} \) rate \( P_h P_0 P_f \)—sufficiently high to prevent any imports, so that all the demand is filled from home supply. Under free trade the price would be \( P_f \), the border price that includes transportation costs to the importing country. The country would then produce \( OH' \) at home and import \( H'M' \). For simplicity we assume that the import supply curve is infinitely elastic at \( P_f \) (the importing country can demand all the com-

Figure 3-10. \textit{Free Trade Is Superior to No Trade}

\[ D' \]

\[ S' \]

\[ 0 \]

\[ H' \]

\[ M' \]

\[ E \]

\[ P_f \]

\[ P_h \]

\[ S \]

\[ E' \]

\[ H \]

\[ P_c \]

\[ P_r \]

\[ \text{Quantity of the importable product} \]

\[ \text{Price} \]

\[ \text{Note: SS'} \text{ is the domestic supply curve. The home price with no trade is } P_h. \text{ A prohibitive tariff would be at the specific rate } P_h P, \text{ or at the ad valorem rate } P_0 P_0 P_f. \]

\[ \text{When trade policy changes from prohibitive protection at } P_h \text{ to free trade at } P_f, \text{ the excess of the gain in consumers' surplus over loss in producers' surplus is } EE'H. \text{ OH' is produced at home, and } H'M' \text{ is imported at } P_f. \]
modity desired at a constant price). When the country moves from no trade to free trade, consumers' surplus increases from \( \Delta \text{P}_{h} \) to \( \Delta \text{P}_{f} \), a gain to consumers of \( \Delta \text{P}_{h} \). Producers' surplus decreases from \( \Delta \text{P}_{s} \) to \( \Delta \text{P}_{f} \), a loss to producers of \( \Delta \text{P}_{s} \). But the gain to consumers is greater than the loss to producers by the amount \( \Delta \text{P}_{h} \). It would therefore be possible for the gainers (the consumers) to compensate the losers (producers) and still be better off in the free-trade situation than in the no-trade situation. Conversely, it would be impossible for the future losers (producers) to bribe the future gainers (consumers) not to “vote” for free trade. If it were politically desired, the government could tax consumers an amount equivalent to \( \Delta \text{P}_{h} \), and taxes could be remitted by this amount for producers. Since potential economic welfare (real income) has increased, the producers could be brought back to their initial income level, while consumers could be left better off after the policy change.

Consider next a change from protection to free trade. In figure 3-11, the price with tariff \( \text{P}_t \) falls to the free-trade price \( \text{P}_f \). Home production would then decline from \( \text{OH} ' \) to \( \text{OH} '' \), while imports would increase from \( \text{H'M}' \) to \( \text{H'M}'' \). The effects would be

- an increase in consumers' surplus: \( \Delta \text{P}_{h} \text{MC}_{f} \)
- a decrease in producers' surplus: \( \Delta \text{P}_{s} \text{HEP}_{f} \)
- a loss in government tariff revenue: \( \Delta \text{P}_{h} \text{HMNT} \).

Gains to consumers exceed losses to producers and the government by the amount of the two shaded triangles \( \text{HTE} \) and \( \text{MNC} \).

Alternatively, we can note that the value of output produced elsewhere in the economy after the tariff is reduced is \( \text{H'EEH'} \) (the area under the supply curve represents the opportunity cost of factors that can be used elsewhere in the

Figure 3-11. Free Trade Is Superior to Restricted Trade

Note: If the tariff \( \text{PP} \) or \( \text{PP/OP} \) is eliminated, the effects are: increase in consumers' surplus, \( \Delta \text{P}_{h} \text{MC}_{f} \); decrease in producers' surplus \( \Delta \text{P}_{s} \text{HEP}_{f} \); loss in government tariff revenue, \( \Delta \text{P}_{h} \text{HMNT} \); net gain, \( \text{HTE} \) (production gain) and \( \text{MNC} \) (consumption gain). An increase in price from the free trade \( \text{P}_{h} \) to price with tariff \( \text{P}_{f} \), would result in the production and consumption “deadweight losses,” \( \text{HTE} \) and \( \text{MNC} \), respectively.
economy). But the imports equivalent to the domestic goods that these resources produced under the tariff now cost only $H'ETH'$. The net production gain to the economy is therefore $HTE$. And the net consumption gain to the economy is $MNC$. Potential welfare has increased. It would be possible, if politically desired, to tax consumers an amount equivalent to the loss in producers' surplus and government revenue, remit taxes to producers so that they are no worse off than before the policy change to free trade, and still leave consumers better off under free trade by the amount of the shaded triangles.

This gain from free trade will, however, be overstated if the resources displaced from domestic production are not absorbed in alternative employment. The value of the released resources that were previously involved in the sheltered domestic production is represented by $H'EHH'$ in figure 3-11. If these resources are not immediately absorbed elsewhere at earnings rates equal to their marginal productivity, then the relevant part of the area $H'EHH'$ that remains idle should be deducted from the two triangles $HTE$ and $MNC$. To be precise, the net direct gain or loss of welfare from reducing the tariff is the present value of the two shaded triangles—determined by discounting the sum over the appropriate time period and at some appropriate interest rate—minus the present value of the loss of productive output during the transitional period when displaced resources are idle. 9

Finally, consider the reverse movement from free trade ($P_f$) to protection ($P_t$). We assume that the tariff does not cause the foreign suppliers to lower their export price—that is, the tariff-imposing country is a small country that cannot affect its terms of trade. Home production would then expand from $OH''$ to $OH'$, while imports would fall from $H'M''$ to $H'M'$. The effects would be

- a decrease in consumers' surplus: $PMCP_f$
- an increase in producers' surplus: $PHEP_f$
- an increase in government tariff revenue: $HMNT$.

The decrease in consumers' surplus is offset in part by the gains to the producers and government. But consumers still lose more than the producers and the government gain. There remain the two shaded triangles $HTE$ and $MNC$, representing the production and consumption “deadweight losses,” respectively.

The production deadweight loss is transferred from consumers to inefficient domestic producers. The consumption deadweight loss means less is consumed at a higher price. It would be impossible for the gainers under protection (producers and government) to compensate the losers (consumers) sufficiently to leave consumers as well off as they were under free trade. If, however, there were a “vote” for protection or free trade, it would “pay” the consumers to bribe producers not to vote for protection (a tariff). Or under free trade, taxes could be imposed on consumers in an amount sufficient to compensate producers and government for lost tariff revenue, while still leaving consumers better off under free trade than protection.

The cost of protection therefore consists of a production cost and a consumption cost. The production cost is the excess real cost of securing $H'H'$ from inefficient home producers instead of from imports. This can be calculated as the “cash cost” or “subsidy-equivalent”—the cash cost to the Treasury of a production subsidy

to the protected industry that would have the same protective effect as the tariff. 10 The subsidy-equivalent of the tariff in figure 3-11 is $P_T^{HT}$. If there were a production subsidy (a negative tariff), designed to obtain the same amount of protection for home industry, the ad valorem subsidy would shift the domestic supply curve down to the new position 55'. This subsidy would raise home output to $OH'$, with a protection effect identical with that of the tariff-equivalent. But there would be no consumption effect because $P_T$ does not rise. And, although there will be a redistribution to producers, the increased revenue to home suppliers comes not from consumers, as in the case of a tariff, but from the Treasury.

The consumption cost under a tariff can be calculated by a compensating variation in income—that is, by the amount consumers' money income would have to increase to leave them as well off after protection as they were in the free-trade situation before prices rose.

From a consideration of the gains and losses in this type of partial analysis, we can conclude that free trade is superior to no trade, 11 and free trade is superior to a tariff.

The virtue of free trade is not only that it makes world production efficient in the technical sense, but it also leads to economic efficiency in the Pareto-efficiency sense of achieving a situation in which it is impossible to make someone better off without making someone worse off. This follows from the fact that, under free trade, consumers face the same prices in both countries; they equate the price ratio to their marginal rates of substitution; and the common marginal rate of domestic transformation in production is equal to the common marginal rate of substitution in consumption. Again, both technical efficiency and exchange efficiency are achieved. Producers are in equilibrium, consumers are in equilibrium, and it would be impossible to make anyone better off without making someone else worse off.

Whether we proceed from no trade to free trade or from a tariff to free trade, in each case the gains from trade are composed of a consumption gain and a production gain. Whether expressed in Ricardo's language as an increase in the "mass of commodities and sum of enjoyments," or as an increase in consumers' surplus, or as an increase in real income, the consumption gain amounts to the substitution of lower-cost imports for higher-cost domestic goods. The production gain arises from the allocation of resources away from the direct production of importables in higher-cost home production to the specialization in the lower-cost production of exportables. The export sector indirectly "produces" the importables on the world market, and this conversion of domestic resources into foreign resources will raise the trading country's real income if the alternative domestic employment of resources used in the export sector would have lower productivity. Moreover, the gains from trade are mutual: every trading nation enjoys a consumption gain and a production gain.

It is clear that the foregoing is the economics of an ideal world. It is not intended as a description of economic reality. On the contrary, its purpose is to provide an ideal reference model against which reality can be compared. It forms a basis for appraising divergences between the optimal and actual. Against a

11. It can also be demonstrated that restricted trade (by tariffs, quotas, or exchange restrictions) is superior to no trade. See M. C. Kemp, "The Gain from International Trade," Economic Journal, vol. 72 (December 1962), pp. 803–19. We are concentrating here, however, on a comparison of free trade and protection.
standard of efficiency, we shall be better able to devise remedial policies to improve upon the actual situation. This view is summarized in the following selection.

3.2 Implications for Economic Welfare

Even if perfect competition were a poorer descriptive tool than it is, students of economics would still want to study it intensively and master its principles. This is so for a reason unconnected with mere description. The competitive model is extremely important in providing a benchmark to appraise the efficiency of an economic system. Russians, Chinese, Indians, as well as Swiss, need to study its analytical principles.

For the most part in science, scholars discuss what is and what will be under this or that situation. The task of positive description is kept as free as is humanly possible from the taint of wishful thinking and ethical concern about what ought to be. Why? Because scientists are cold-blooded robots? No. because experience shows that a more accurate job of positive description will be achieved if one tries to be objective. (Experience also shows that, try as we may, we humans never succeed in separating completely the objective and subjective aspects of a discipline. Indeed, the very choice of what scientists choose to measure, and the perspective from which they observe and measure it, and the reactions the observer produces in that which is observed—all these make the distinction between is and ought, between objective and subjective issues, at bottom a matter of degree rather than kind.)

The citizenry, unlike the specialist, is in the end most interested in problems of norms, of what ought to be done, of policy rather than mere description. That citizenry is best served by the scientist who can give it the most accurate description of what is relevant, and of what the consequences of different policy actions will be.

Very briefly then, in this epilogue, let us again devote a few lines to implications for economic welfare that have been developed in [this chapter].

Efficiency Optimality

First, we have seen repeatedly that a regime of perfectly competitive pricing would have certain efficiency properties if conditions were really present for maintaining perfect competition. Adam Smith, in his talk about an Invisible Hand which leads the selfish actions of individuals toward so harmonious a final result, did have some point. Smith never could state or prove exactly what that point was, but modern economics can state this property of ideal competitive pricing:

Under perfectly perfect competition, where all prices end up equal to all marginal costs, where all factor-prices end up equal to values of marginal-products and all total costs are minimized, where the genuine desires and well-being of individuals are all represented by their marginal utilities as expressed in their dollar voting—then the resulting equilibrium has the efficiency property that "you can't make any one man better off without hurting some other man."

What does this mean exactly? It means that a planner could not come along with a slide rule and find a solution, different from the laissez-faire one, which could improve the welfare of everyone. 12

Arbitrary Distribution of Dollar Votes

What does competitive efficiency not mean? It does not mean that actual laissez-faire, with the imperfections of competition that will go with it, leads to efficiency or necessarily even to a close approximation to efficiency. And inefficiency aside, it does not mean that the people who are deemed by various religious or ethical observers to be most worthy, most deserving, or most needy will necessarily get their ethically best share of goods and services.

Laissez-faire perfect competition could lead to starving cripples; to malnourished children who grow up to produce malnourished children; to perpetuation of Lorenz curves of great inequality of incomes and wealth for generations or forever. Or, if the initial distribution of dollar-wealth votes, genetic abilities, early conditioning, and training happened to be appropriate, perfect competition might lead to a rather egalitarian society characterized by uniformity greater than might please many an aristocratic, ethical tra-

12. If A is the competitive equilibrium, there is no point B the planner could devise which could be approved over A by a unanimous vote. Some would be hurt in going to B, some might gain; but the gainers could never find it worthwhile to give big enough bribes to win the losers over to approving the move to B. We've already encountered this notion of "Pareto optimality," named after Pareto's work at the turn of the century as elucidated in our time by Abram Bergson of Harvard.
dition. Or more likely, lead to inequality deemed too glaring.

In short, Adam Smith had no right to assert that an Invisible Hand successfully channels individuals who selfishly seek their own interest into promoting the "public interest"—as these last two words might be defined by a variety of prominent ethical and religious notions of what constitutes the welfare of a nation. Smith has proved nothing of this kind, nor has any economist since 1776.

Redefining the Invisible Hand Doctrine

If Smith were alive today, he would agree with all this; and one ventures the guess, from his biography, that he would probably reword his doctrine pretty much along the following lines.

1. Only if abilities and dollar-wealth votes were originally distributed in "an ethically optimal" manner—and kept so distributed by nondistorting, nonmarket interventions—could even perfectly competitive pricing be counted on (a) to produce an efficient configuration of production on society's production-possibility frontier (and not inside it), and (b) to give people what they really deem is best for them, in accordance with their dollar votes that now reflect equally significant social utility. But, if laissez faire were abandoned in favor of an ethically proper distribution of wealth and opportunity, then perfect-competition equilibrium could be used as an instrument to attain optimally efficient and equitable organization of society.

2. Admittedly, the demands of people in the marketplace sometimes do not reflect their true well-beings as these would be interpreted by even the most tolerant and individualistic of ethical observers. (Examples: A dope addict's craving for heroin even at the expense of his food or his children's food; a child's desire for a seventh lollipop; a diabetic's craving for sweets; a spendthrift's mortgaging his house for an advertised sports car.) As a pragmatist, were Smith alive today, he would say, "People are entitled to make their own mistakes in many matters, but it is arrogant to think that anyone who is not a minor or a certifiable mental incompetent is in every respect a sovereign will; as men, conscious that we were born not perfect and as inhabitants of a post-Freudian world, we shall sparingly and for good cause want (with due process) democratically to place restrictions on our own behavior. So again, laissez-faire perfect competition would not inevitably be the ideal."

Then, after emphasizing the need for tolerance and the virtues of freedom, Smith might strike a more technical note.

3. Where there are monopolistic imperfections that produce deviations from ideal competitive marginal-cost pricing—and in situations of strong increasing returns and decreasing costs, such deviations are practically inevitable—of course there is a prima facie case that laissez-faire pricing is not efficient. Public scrutiny, to see whether democratic controls would make the situation better or worse, has to be presumed in such quasi-monopoly situations.

4. Finally, Smith would add that wherever there are "external economies and diseconomies," there is a prima facie case for study to see whether zoning laws, taxes or subsidies, and government expenditure and regulation should be initiated in some degree. Where the checks and balances of perfect competition are not operative, the Darwinian struggle for existence is not led by an Invisible Hand to any kind of optimum. The creative role of government in economic life is vast and inescapable in an interdependent, crowded world.

Market Imperfections

We now descend from the ideal world to the actual and proceed to redefine the Invisible Hand doctrine, as suggested in the foregoing selections. We must now allow for cases of market failure and consider what types of corrective interventions the government might introduce to improve upon the actual operation of markets.

Market failure can exist in various forms: (1) the market does not function properly—the case of market imperfections; (2) the market result is incorrect—the case of externalities; (3) no market exists for the relevant activity—the case of public goods; and (4) the market yields undesirable results in terms of objectives other than resource allocation. In these cases of market failure, market prices do not exist, or do not reflect the true value, or are irrelevant.

A state of underdevelopment is in large measure synonymous with pervasive market imperfections. A variety of market imperfections can reduce the efficiency of private market performance. The properties of efficiency in resource allocation
and Pareto optimality of distribution, which characterize a competitive equilibrium, depend on the existence of a competitive set of markets. But an underdeveloped economy is to some extent an "empty economy" with an incomplete set of markets. Moreover, unlike the perfectly competitive markets we have previously postulated, a market may be deficient in the provision of information, or subject to lags in adjustment, or characterized by insufficient competition among firms.

If information is inadequate or inaccurate, market signalling will not be read or correctly interpreted by agents in the market. A certain resource allocation will occur, but it will differ from the efficient allocation based on adequate and accurate information. It can then be argued that it is in the public interest to have the government correct these informational deficiencies. The scope for the provision of public information is wide in a developing economy—from agricultural extension services to technical assistance to information on world market opportunities.

Time lags also impede the efficient working of the market. While resources might respond to market signals, they may do so only after an unduly long delay. To accelerate the supply response to a rise in price, for example, the government may want to undertake policies to stimulate supply. This is especially relevant for overcoming factor immobility, via training and educational measures and other policies designed to reduce barriers to entry in an occupation or industry. After acquiring information about a market opportunity, the economic agent must be willing and able to react on this information. To fulfill these conditions more speedily, the government in a developing country may have to undertake policies to overcome the fragmentation or segmentation of markets and improve the operation of markets. And yet while these market imperfections may prompt government intervention, it is necessary to weigh the costs of arbitrary direct administrative controls and to be sure that they do not exacerbate an already fragmented economy. This is emphasized in Selection 3.3.

### 3.3 The Fragmented Economy

While economists can usefully divide their labor as monetary theorists, tax experts, foreign trade specialists, project evaluators, and so on, a unified view of the development process is a great analytical convenience. Why is public intervention so pervasive and generally so unsuccessful? Intervention is usually prompted by the perception—sometimes correct—that a particular market is functioning badly, so that authorities feel pressed to "do something." An infant textile firm is helped by a tariff; or the price of an agricultural product may be raised to permit farmers to use a new fertilizer-intensive technology; or a tax exemption may be granted to a foreign firm for automobile assembly. This pressure for public intervention is the result of severe fragmentation in the underdeveloped economy.

The economy is "fragmented" in the sense that firms and households are so isolated that they face different effective prices for land, labor, capital, and produced commodities and do not have access to the same technologies. Authorities then cannot presume that socially profitable investment opportunities will be taken up by the private sector, because prevailing prices need not reflect true economic scarcity—at least not for large segments of the population. There is historical justification for this view in the nineteenth and early twentieth centuries. In Asia, Latin America, and Africa, primary commodity export enclaves were controlled by foreigners, and much of the general population remained outside of the market economy. Indigenous entrepreneurs had limited access to capital, no means of acquiring advanced technologies, and little skilled labor.

Newly independent governments quite properly felt compelled to act as agents of change to offset

economic and political colonialism. In the past twenty or thirty years, poor countries have succeeded in introducing some new industrial activities—particularly the manufacture of goods previously imported—and in mobilizing some domestic factors of production. Their governments chose to do so, however, by manipulating commodity prices in a variety of ways and by intervening directly to help some individuals or sectors of the economy at the expense of others.

Consider the extraordinary lengths to which import tariffs have been used in Latin America, with rates of several hundred percent on some goods and absolute prohibitions on the import of others, while still others enter freely. The situation on the Indian subcontinent is no different. Price and quantity controls on foreign trade and domestic commerce make licensing and rationing commonplace. Byzantine patterns of industrial taxes and subsidies complicate government budget making. Consequently, the market mechanism has become no better, and perhaps even worse, as an indicator of social advantage.

Modern fragmentation, therefore, has been largely the result of government policy and goes beyond the old distinction between the export enclave and the traditional subsistence sector. One manifestation is the often-noted existence of small household enterprises and large corporate firms—all producing similar products with different factor proportions and very different levels of technological efficiency. Continuing mechanization on farms and in factories in the presence of heavy rural and urban unemployment is another. Excess plant and equipment with underutilized capacity are commonly found in economies that are reputed to be short of capital and that do suffer from specific bottlenecks. In rural areas, tiny landholdings may be split up into small noncontiguous parcels, with inadequate incentives for agricultural land improvements.

While tangible land and capital are badly used, fragmentation in the growth and use of human capital can be more serious and no less visible. Learning-by-doing and on-the-job training in the "organized" economy are confined to narrow enclaves—export-oriented in the past but now increasingly inward-oriented toward "modern" manufacturing—whose employment growth may be less than the growth in general population. Unemployment among the highly educated coexists with severe shortages in some labor skills.

Indigenous entrepreneurship is narrowly based and is supported by heavy government subsidy. Tariff protection, import licenses, tax concessions, and low-cost bank finance commonly go to small urban elites and create great income inequality between the wealthy few and the poverty-stricken many. This income inequality has failed to induce high rates of saving in the classical manner, but governments remain reluctant to reduce the disposable income of well-to-do investors whose unique access to investment opportunities is guaranteed by the web of official controls and by the endemic fragmentation.

**Liberalization and the Capital Market**

How does one begin to loosen the Gordian knot? The incredibly complex distortions in commodity prices now prevailing are the unplanned macroeconomic outcome of specific microeconomic interventions. But substantial fragmentation in the markets for land, labor, and capital provided the initial motivation for public authorities to "do something" and continues to pressure governments to intervene. Thus an explicit policy for improving the operation of factor markets is necessary to persuade authorities to cease intervening in commodity markets. Carefully considered liberalization in all sectors can then move forward—not merely as a reaction to the more obvious mistakes of the immediate past, but in ways that allay legitimate fears of pure laissez-faire.

However, the knot needs to be loosened further. To say that there are "imperfections in factor markets" is distressingly vague and often signals the end of formal economic analysis. But further systematic inquiry can proceed if the neoclassical approach of treating labor, land, and capital symmetrically as primary factors is dropped. It is hypothesized here that fragmentation in the capital market—endemic in the underdeveloped environment without carefully considered public policy—causes the misuse of labor and land, suppresses entrepreneurial development, and condemns important sectors of the economy to inferior technologies. Thus appropriate policy in the domestic capital market is the key to general liberalization, and particularly to the withdrawal of unwise public intervention from commodity markets.

A major source of market imperfections in a developing country is the inability of new firms to enter the market and instill greater competition. Instead of unrestricted competition, the market is more apt to be characterized by a monopoly or a few sellers that dominate the market. The entry of newcomers is frequently prevented or limited by natural obstacles, such as economies of scale and the uninformed nature of the market, or by the deliberate action of a group of
The hallmark of a competitive market is that every participant in it is a price-taker. That is, everyone in it takes it for granted that he cannot affect the prices at which he buys or sells, so does the best he can in the light of those prices. For this situation to hold there must be a large number of suppliers in the market, tolerably equal in size, so that no one firm can have an appreciable effect on the price or aggregate supply of the commodity, and so that no one firm's actions significantly influence the fortunes of any other firm. It must also be true that consumers do not care much, if at all, which firm they buy from, so that they respond readily to small differences in the prices charged by different firms if any should arise. Some examples of competitive markets are women's clothing in manufacturing, coal in mining, and wheat in agriculture. In all those industries the firms are numerous and small, and the products of one firm are virtually indistinguishable from those of any other.

But modern methods of mass production and mass marketing tend to be incompatible with competition, and it is easy to see why this is so. Consider, first, modern methods of production. They tend to require large amounts of fixed capital equipment, and to be characterized by decreasing long-run costs. That is to say, a large plant operating at an efficient volume of output will have lower average costs than a small plant operating at its most efficient level of output.

When these technological conditions obtain, even an industry that starts out competitive will soon degenerate into some other market form, because the firms in it will tend to grow in order to reduce their costs of production by operating larger, more efficient plants. But they will not all grow synchronously. Historical accidents will guarantee that some of the firms will build large, modern plants before the others, and these will be under strong pressure to cut their prices in order to attain sales volumes at which the large plants can operate efficiently and cover their enhanced fixed cost. Furthermore, by virtue of their increased efficiency, the large firms will be able to operate profitably at prices below the average costs of their smaller competitors. This will produce an unstable situation. One by one, the smaller firms will be driven out of the market, forced to bequeath their erstwhile customers to the larger firms—which therefore will grow larger still. Eventually only a few firms will remain, all of them large enough to reap the full advantages of economies of scale.

When this state of affairs is reached, true competition cannot survive. The exact outcome depends on many factors, including especially the proportion of the market that is served by a firm large enough to be fully efficient. We saw that frequently there is room for only two or three fully efficient firms, even in a large national market. Now, General Electric and Westinghouse cannot pretend that they are participants in a competitive market; both know that between the two of them they control the price and volume of output of most of the products they sell, and this knowledge cannot help but affect their behavior. They constitute an oligopoly. To repeat our definition:

An oligopoly is a market in which there are a few firms, each of which recognizes that its actions have a significant impact on the price and supply of the commodity.

The process of consolidation that we have just described can go even further if economies of scale continue to accrue (i.e., if the long-run average-cost curve continues to decline), until a single firm is so large that it supplies the entire market, or at least the great bulk of it. In that case only one firm will remain, and a monopoly will be established.

A monopoly is a market in which all or virtually all of the commodity is provided by a single seller.

Besides, as we shall see, oligopolists have strong incentives to merge their firms into a monopoly, which from their point of view is the more profitable market form.

An extreme form of economies of scale is a so-called natural monopoly, exemplified by local electric power companies, telephone companies, and other public utilities. There the technical advantages of supply by one large firm are so commanding that a single firm is licensed by the government to serve the entire market. In return for this exclusive license, or franchise, the firm submits to control by regulatory authorities which endeavor to assure adequate service at reasonable prices on the part of firms that are insulated from any competitive pressures. The regulatory agencies, in effect, try to prevent the mo-

nopolies from behaving like monopolies. We shall shortly see why this is important.

To summarize briefly: the advantages of mass production, or economies of scale, engender oligopoly and, in extreme form, monopoly; the advantages of mass marketing reinforce, and in many cases replace, the effects of economies of scale.

Monopoly

Everyone envies the happy monopolist. The competitive firm has to sell at the market price, and squeeze out what profits it can by holding production costs down to their very minimum—but even then is likely to have its profits competed away. The oligopolist has formidable rivals to contend with. But the monopolist chooses his own prices and has no serious rivals; indeed, he has practically nothing to worry about except the Anti-Trust Division, or perhaps a regulatory commission. How should he conduct himself to take advantage of this unparalleled felicity?

The answer is basically a familiar one by now, but we have to introduce a new concept to express it clearly. The essential peculiarity of the monopolist is that his firm's individual demand curve is identical with the demand curve for his product. The monopolist has the privilege of setting the price of his product, but then he can sell only the quantity corresponding to that price on his product's demand curve. If he raises the price he must be content to sell less; if he wants to sell more he must lower the price. So his essential decision problem is to find the most profitable price-quantity pair on his demand curve. He does this by the now-familiar process of equating his margins—by equating the marginal cost of his product to the increase in his gross sales revenue that would result from selling one more unit. This is the concept we need, the concept of marginal revenue.

Marginal revenue is the increase in the gross value of sales that would result from selling one more unit.

The distinctive feature of a monopoly is that a monopolist's marginal revenue is not equal to his price, as is a competitive firm's. For the monopolist to sell an additional unit he must reduce the price on all units sold, in accordance with the slope of his demand curve. Hence the amount he realizes by selling an additional unit is less than the price he sells it for; reduction in the price of his previous sales has to be subtracted off. For a monopolist, though not for a competitor, then, we must distinguish between price and marginal revenue.

It should be pretty evident why it is most profitable for a monopolist to select the volume of sales at which marginal revenue equals marginal cost. If marginal revenue is greater than marginal cost, selling one more unit will contribute more to his sales revenues than producing it will add to his expenses. It will enhance his profits. On the other hand, if marginal revenue is less than marginal cost, he is losing money on the last unit he produces, and perhaps on others. But when marginal revenue is equal to marginal cost, he cannot increase his profits either by reducing or increasing his levels of output and sales.

From a competitive firm's point of view, the demand curve for his product is too remote to be of interest; no conceivable increase in his level of sales could have a visible effect on the total sales of the industry or on the market price. But a monopolist realizes that he can depress the price by his own activities, and this makes increases in output less attractive to him than to a competitive firm.

To see the force of all this, suppose that a monopolist can sell 800 units a month at a price of $13 per unit and that his demand curve shows that he has to reduce his price by 5¢ in order to sell 801 units a month. Then, with an output of 800 units a month, his total value of sales is $10,400; with an output of 801 units a month his total value of sales is $10,407.99. Note that his marginal revenue is about $8, which is substantially less than the price of $13 which he is charging.

Whether the monopolist will want to increase his monthly sales from 800 to 801 depends on the comparison of this marginal revenue with his marginal cost. If his marginal cost at an output of 800 a month is less than $8 per unit, it will be profitable for him to increase his production. But if his marginal cost is more than $8 a unit, even though it is less than his price of $13 a unit, he will want to reduce his output to 799 a month (which would permit him to raise his price about 3¢ per unit), or even further. He will be content with sales of 800 a month only if his marginal cost is just about $8 a unit. In general:

A monopolist will choose a price-quantity combination on his demand curve at which his marginal revenue is equal to his marginal cost.

This is the salient difference between competition and monopoly: the competitor and the monopolist both produce outputs such that marginal revenue equals marginal cost, but a competitor's marginal revenue is equal to his price, while a monopolist's is...
inevitably less. We shall soon see the consequence of this.

### A GRAPHIC REPRESENTATION OF MONOPOLY

The behavior of a monopolist can be analyzed with the help of a convenient diagram. To help construct it, imagine a monopolist whose cost data are the same as those of the competitive firm. (For convenience, those cost data are reproduced in table 3-4 and graphed in figure 3-12.) A monopolist does not confront a market price, but he does have to consider the demand curve for his product, for this shows how many units he can sell at each price that he might select. This curve also is shown in figure 3-12.

The worthwhileness of an additional sale per month to a monopolist does not depend upon its price but upon the marginal revenue that it yields, and we already know that this will be less than the price. The marginal revenue corresponding to each level of sales is also shown in the figure.

The marginal-revenue and marginal-cost curves intersect at an output of 800 units a month. (The average-cost curve is irrelevant for the moment.) By our previous reasoning, this is the most profitable output for the monopolist, so that he should charge the corresponding price on the demand curve, or $13. If he should charge a higher price (sell fewer units) he would be declining to produce some units that would add more to his revenue than to his costs, because for them, marginal revenue would exceed marginal cost. If he should charge a lower price, he would be selling some units that cost him more than their contribution to his sales revenue. Only $13 is just right.

How about his profit at this most profitable level of output? Profit arises from the excess of price over average cost. It so happens, in this instance, that the average and marginal costs are equal at the most profitable level of output. (We shall see shortly why the data were chosen to work out that way.) Therefore, marginal revenue is equal to average cost, and since price is necessarily greater than marginal revenue, there is a positive profit—the monopolist's profit.

But, we should emphasize, the data didn't have to work out that way. The most profitable level of output, determined by the intersection of the marginal-cost and revenue curves, might have occurred in the rising portion of the average-cost curve. In that case, marginal cost would have been greater than average cost, and price, being greater than marginal revenue (= marginal cost), would have been greater still. Finally, the most profitable level of output might have occurred in the falling portion of the average-cost curve. Then marginal cost and marginal revenue would be less than average cost. But the price, which exceeds the marginal revenue, could still be greater than the average cost, leaving a positive profit.

### MONOPOLY IN THE LONG RUN

In the short run, then, the monopolist with a given fixed plant operates at the level of output at which marginal revenue and marginal cost are equal. His long-run adjustments are then very much like those of a competitive firm. Having attained the most profitable level of output in the short run, the monopolist will enlarge, reduce, or modify his plant if by doing so he can reduce the average cost of his current level of output. In the long run, therefore, he will have the plant in which his current output can be produced as cheaply as possible.

If the monopolist's technology is one with constant long-run costs, then just as in a competitive industry,

<table>
<thead>
<tr>
<th>Monthly output (units)</th>
<th>Average cost ($/Unit)</th>
<th>Marginal cost ($)</th>
<th>Price ($)</th>
<th>Marginal revenue ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>19.00</td>
<td>—</td>
<td>17.375</td>
<td>16.74375</td>
</tr>
<tr>
<td>200</td>
<td>13.00</td>
<td>7.00</td>
<td>16.750</td>
<td>15.49375</td>
</tr>
<tr>
<td>300</td>
<td>10.67</td>
<td>6.00</td>
<td>16.125</td>
<td>14.24375</td>
</tr>
<tr>
<td>400</td>
<td>9.50</td>
<td>6.00</td>
<td>15.500</td>
<td>12.99375</td>
</tr>
<tr>
<td>500</td>
<td>8.80</td>
<td>6.00</td>
<td>14.875</td>
<td>11.74375</td>
</tr>
<tr>
<td>600</td>
<td>8.42</td>
<td>6.50</td>
<td>14.250</td>
<td>10.49375</td>
</tr>
<tr>
<td>700</td>
<td>8.21</td>
<td>7.00</td>
<td>13.625</td>
<td>9.24375</td>
</tr>
<tr>
<td>800</td>
<td>8.19</td>
<td>8.00</td>
<td>13.000</td>
<td>7.99375</td>
</tr>
<tr>
<td>900</td>
<td>8.33</td>
<td>9.50</td>
<td>12.375</td>
<td>6.74375</td>
</tr>
<tr>
<td>1,000</td>
<td>8.65</td>
<td>11.50</td>
<td>11.750</td>
<td>5.49375</td>
</tr>
<tr>
<td>1,100</td>
<td>9.14</td>
<td>14.00</td>
<td>11.125</td>
<td>4.24375</td>
</tr>
</tbody>
</table>
when he has the most appropriate plant for his current output, he will also be using that plant at its most efficient level: the bottom of its average-cost curve. This is the situation depicted in figure 3-12, which accordingly shows the long-run equilibrium of a monopolist with constant long-run costs. By contrast with the competitive case, in which profit is competed away in the long run, a substantial margin of profit remains for the monopolist.

We know, however, that monopoly typically tends to arise in a different sort of industry, one characterized by decreasing long-run costs. The situation shown in figure 3-13 is therefore much more characteristic of monopolies. In that figure, the curve labeled AC is a long-run, average-cost curve—that is, a curve showing the average cost for each level of output when produced in the plant in which that cost is lowest. (Note that it declines throughout its range.) The corresponding marginal-cost curve is also shown; it is always below the average cost curve. (Why?) This diagram makes it clear that a monopolist cannot be blamed for behaving like a monopolist. Since his marginal cost is always lower than his average cost, if he were to charge a price equal to marginal cost, as a competitor does, he would be losing money. No matter how economists may extoll the virtues of marginal-cost pricing (i.e., charging for each unit of a product the addition to costs attributable to it), the monopolist cannot afford to follow that policy.

Demand and marginal-revenue curves are also shown in figure 3-13. The long-run equilibrium for this monopolist is to produce about 960 units a month for a marginal cost of 68¢ and a price of $10.40. A regulatory agency might require the monopolist to charge a price equal to average cost, interpreting cost to include a reasonable return on invested capital; and this is often done. In this instance, however, extrapolating the data in figure 3-13 shows that such a regulation would result in an output of 1,850 units, a price of $1.50, and a marginal cost of 54¢. Although the monopolistic restriction of output would be greatly reduced, the principle of marginal-cost pricing would still be violated.

**MONOPOLY AND ELASTICITY OF DEMAND**

What makes monopoly possible is the efficiency of large-scale operation; what makes it worthwhile is the slope of the demand curve. For remember: the more nearly horizontal a monopolist's demand curve is, the closer together are his demand and marginal-revenue curves, and the smaller is the discrepancy between price and marginal cost. If a monopolist could have an absolutely horizontal demand curve, he would behave just like a competitor. The difference between competitive and monopolistic behavior therefore depends on the slope of the demand curve. Perhaps we had better see in a bit more detail how this slope influences the monopolist.

For this purpose, the most significant characteristic of a demand curve is its elasticity, a very natural measure of how responsive demand is to changes in price.

The elasticity of demand for a commodity is the percentage reduction in the volume of sales that would be induced by a 1 percent increase in price.

To illustrate this concept, let us compute the elasticity of demand for the monopolized product of table
3-4 when 800 units are being sold. The price is then $13, and a 1 percent increase would raise it to $13.13. From the formula for the demand curve we can compute that 779 units would be sold if this price were charged. This is a fall of 21 units or 2.6 percent, which accordingly is the elasticity.

The critical value of elasticity is unity, or one. For if elasticity = 1, then a 1 percent increase in price will cause a 1 percent decrease in the volume of sales. These two changes will just about offset each other, leaving the total value of sales unchanged. But with 1 percent fewer units to be produced, total variable costs would be diminished by such a change. It follows that if a monopolist should find himself producing an output for which the elasticity of demand is unity or lower, he would restrict his output and increase his price. As he does so, his elasticity of demand will increase, generally speaking (but not inevitably). This is illustrated in table 3-5, which is derived from table 3-4, by the computation just explained. The monopolist will stop raising his price, as we already know, when he has restricted output to the point where marginal revenue and marginal cost are equal.

When this equality is attained, the percentage markup of price over marginal cost, from which the monopolist's profit is derived, is equal to the reciprocal of the elasticity of demand. If the elasticity of demand is high, this percentage markup is small; if demand is inelastic it is large. The profitability of a monopoly depends, therefore, directly on the elasticity of the demand curve.

The fundamental objection to monopoly is that by holding down production to the level at which marginal cost equals his marginal revenue, the monopolist wastes economic resources. It is important to see why this is so. When the output is 800 units a month, the marginal cost is $8 per unit. This means that an 801st unit could be produced by purchasing and using labor services, raw materials, power, and other ingredients worth, in all, $8. The fact that the monopolist would have to pay $8 for these ingredients indicates that other firms—especially including competitive ones—are willing to pay that much for them; otherwise the price would be different. This indeed is the true meaning of prices: what one man has to pay for something is what another one would be willing to pay for it.

Now, why would a competitive firm be willing to pay $8 for these resources? Because it could use them to produce something worth $8 to its customers, thereby adding $8 to its total sales. (Remember that in a competitive industry price equals marginal cost, so that a product that adds $x to costs will sell for $x.) We have encountered again the notion of opportunity costs—the idea that the real cost of anything is the value of the things that could have been produced with the resources it consumes. The opportunity cost of the monopolist's unproduced 801st unit, for which customers are willing to pay about $12.99, is only $8, and herein lies the waste. Since the customers would be willing to pay $12.99 for the unavailable 801st unit, they clearly prefer it to the $8 worth of alternatives that are produced in its stead. By declining to produce the 801st unit, the monopolist has compelled the economy to use certain resources to produce something worth only $8 when he could have used those resources to produce something worth $12.99. The monopolist has barred some resources from their most productive employment.

We can look at this same distortion in another way: by pegging the price at $13, the monopolist has

<table>
<thead>
<tr>
<th>Monthly output (units)</th>
<th>Price ($)</th>
<th>Marginal revenue ($)</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>17.38</td>
<td>16.744</td>
<td>27.8</td>
</tr>
<tr>
<td>200</td>
<td>16.75</td>
<td>15.494</td>
<td>13.4</td>
</tr>
<tr>
<td>300</td>
<td>16.12</td>
<td>14.244</td>
<td>8.6</td>
</tr>
<tr>
<td>400</td>
<td>15.50</td>
<td>12.994</td>
<td>6.2</td>
</tr>
<tr>
<td>500</td>
<td>14.88</td>
<td>11.744</td>
<td>4.8</td>
</tr>
<tr>
<td>600</td>
<td>14.25</td>
<td>10.494</td>
<td>3.8</td>
</tr>
<tr>
<td>700</td>
<td>13.62</td>
<td>9.244</td>
<td>3.1</td>
</tr>
<tr>
<td>800</td>
<td>13.00</td>
<td>7.994</td>
<td>2.6</td>
</tr>
<tr>
<td>900</td>
<td>12.38</td>
<td>6.744</td>
<td>2.2</td>
</tr>
<tr>
<td>1,000</td>
<td>11.75</td>
<td>5.494</td>
<td>1.9</td>
</tr>
<tr>
<td>1,100</td>
<td>11.12</td>
<td>4.244</td>
<td>1.6</td>
</tr>
<tr>
<td>1,440</td>
<td>9.00</td>
<td>-0.006</td>
<td>1.0</td>
</tr>
</tbody>
</table>
put a false signal into the price system. From the standpoint of his customers, the opportunity cost of his product is $13 because that is the value of the other things they have to forgo in order to obtain a unit of it; from the standpoint of the economy the opportunity cost is $8. Consumers therefore unduly restrict their use of the monopolized product. Perhaps they use an inferior substitute for it that costs $11 to produce, thereby consuming $11 worth of resources for a purpose that could be served better by $8 worth. Consumers are misled, from the point of view of making the best use of the resources available to the economy, into buying too little of the monopolized product, and too much of others.

For the foregoing reasons, among others, unregulated monopoly is generally considered to be intolerable. But in the public service industries, especially, the economies of scale are so striking that it seems almost inconceivable to have several firms serving a single area or route. Consequently, in these industries it is usual to permit a monopoly subject to government supervision.

When competition is restricted, inefficiency results. Once the seller restricts output to raise price and maximize profits, price becomes higher than marginal cost, underutilization of plant capacity occurs, and output is less than ideal. It will be remembered that the economic efficiency of unrestricted competition consisted in price equalling marginal cost, the marginal valuation of the consumer equalling the marginal cost of production, and the maximization of the sum of consumers’ and producers’ surpluses. Selection 3.5 illustrates the loss due to economic inefficiency.

3.5 Economic Inefficiency

We are now ready to show graphically the loss due to economic inefficiency. Imperfect competition and monopoly enable either the sellers or the buyers to restrict the volume of transaction in order to gain a price advantage over the person or persons on the other side of the market; and while most of the other party’s (or parties’) loss is their gain, a small part of it is nobody’s gain but a net loss to society. This is shown diagrammatically in figure 3-14(b) for the case when the sellers are the price makers, able to raise price and restrict output. The producers’ surplus is bigger and the consumers’ smaller than in the perfectly competitive case (shown in figure 3-14(a); representing the exploitation of consumers by producers; the net loss to society, caused by the economic inefficiency of imperfect competition, is represented by the little triangle shaded in figure 3-14(b). The height of the triangle is the producers’ profit margin; its width is the difference between what output is and what it would be under perfect competition. The area of the triangle therefore is relatively easy to calculate; and such calculations are the basis of all empirical estimates of the value of loss due to economic inefficiency.

It is obvious from the diagram that the net loss is likely to be small compared to the redistribution of income between buyers and sellers. This expectation is fully borne out by empirical estimates. All quantitative estimates of the losses due to various restraints on competition have shown these to be a very small proportion of the total value of the output or transactions affected. Furthermore, a summary comparison of estimates of losses due to economic as against technological inefficiency has shown the former to be very much smaller than the latter.

These findings strongly suggest that of all the bad effects of imperfect competition, the most serious ones are likely to be those that stem from restraints on entry and the consequent elimination of that competitive pressure which keeps everybody-managers, engineers, workers—on the top of their form. The economic inefficiencies due to prices’ transmitting wrong signals in the imperfect market seem, on this evidence, much less important. If free entry is, indeed, the crucial factor to ensure the efficiencies that matter, then the difference in performance between free and restricted competition must also be crucial.

Just as market imperfections in the form of oligopoly and monopoly are important in the domestic economy, so too are they significant in the international economy. The transnational corporation may give rise to market imperfections,
especially in the practice of transfer pricing. This is explained in Selection 3.6. International commodity agreements also introduce some noncompetitive features, as noted in Selection 3.7.

3.6 Transfer Pricing and Developing Countries

This paper intends to review some of the main problems which arise for developing host countries in the investigation of transfer pricing of commodity trade by transnational corporations (TNC) in manufacturing industry. It concentrates on the pricing of intra-firm trade (i.e., trade between associated units of a company operating in several countries), but part of the analysis will also be relevant to the pricing of inter-firm trade (between unrelated parties) by TNC in developing countries.

The available evidence indicates, as we shall see, that the 'effectiveness' of transfer-pricing for TNC (the extent to which profits can be clandestinely moved by changing intra-firm prices as compared to declaring profits) varies widely from industry to industry and from one firm to another. This is worth taking into account for policy purposes: on the one hand, it makes it more difficult for host governments to formulate general rules and procedures for checking price manipulation, but, on the other, it provides an opportunity to narrow the field which they need to monitor in order to check the largest potential abuse. For governments with limited administrative resources, this is clearly an important consideration.

Three sets of factors account for the uneven incidence of transfer pricing. First, the trade component of TNC production itself varies according to industry. Second, the extent of intra-firm as a proportion of total trade by TNC differs widely, both according to industry, and also possibly according to the national origin of the firms involved. Third, the possibilities of manipulating intra-firm prices to shift revenues also vary from case to case.

Let us assume that a host government in a less developed country has access to all the production cost, fixed cost, intra-firm prices and profit data it wants on a TNC with operations in its jurisdiction. Let us assume also that the firm concerned is in a high technology industry, with heavy R and D expenses, a lot of intra-firm trade in firm-specific products and with considerable market power in final product markets. Can it then proceed in a straightforward manner to assign 'correct' transfer prices on the goods which enter or leave its country?

Unfortunately not. A 'correct' transfer price must be one which is fair in some sense to both the government and the TNC concerned, and it must be compatible with the continuation of desirable technical progress in that industry. It must, in other words, be a price which provides a just reward for investments in R and D by that TNC, bearing in mind the riskiness of innovation in the industry, but not one which permits unjustified monopoly profits or socially wasteful expenditures on building up monopoly power.

Figure 3-14. Losses Due to Economic Inefficiency

The complexities of the issues can be illustrated with the example of the pharmaceutical industry, the prime target of transfer-pricing investigations.

The transnational pharmaceutical industry is heavily dependent for its success on innovation and brand-name promotion. TNC in this industry indulge in a great deal of intra-firm trade, and it is well known that the prices charged for intermediates, in developed and developing countries, are very much higher than the direct costs of production. The fact that TNC can maintain a dominant position in world markets despite charging prices, in intermediate and final markets, much higher than smaller firms which imitate their products depends partly on their technological leadership (legally maintained for certain periods by the institution of the patent system) and partly on brand preference created by extensive promotion.

The bulk of investigations of the pharmaceutical industry in developing countries have defined the correct transfer as the price at which they can obtain the chemical from another supplier: its 'opportunity cost' to the country. For new drugs, the alternative supplier is an imitator from a non-patent-observing country; for older drugs, it may be competitors (of all sizes, but generally small firms which sell unbranded, or generic, drugs) from developed and less developed countries. The TNC concerned defend their admittedly higher prices by referring to the expense and riskiness of the R and D involved (for new drugs) and by the better quality (and also contribution to the R and D effort) of older drugs.

It is clear that both sides have a point. The TNC are right in that the price charged by an imitator does not represent a correct transfer price for an innovating firm: if this price were charged on a large enough scale, the drug itself would not be brought into existence. It is not a price which is compatible with (presumably socially desirable) innovation. The developing countries are right to the extent that the transfer price set by the TNC may be much higher than is required to provide a just return to innovation: it may incorporate returns to market power created by promotion and the abuse of patent protection. The matter is further complicated by such factors as: (a) particular successful drugs must bear the burden of financing R and D on hundreds of others which fail; (b) particular host countries may be interested only in some of the innovations of a TNC and so may not wish to contribute to R and D on others; (c) developing countries may wish to contribute at a lower rate than developed countries for drugs (like psychotropic or cancer drugs) which are intended primarily for rich markets; and (d) developing countries may feel that a large part of the promotional expenditure (which usually ranges from 15–25% of the price of drugs) that is required to commercialize on innovations is wasteful, and serves only to perpetrate barriers to entry of competition.

It should be clear that, in such a situation, it is impossible to define a transfer price which is correct in an objective sense. Reference to prices charged by the TNC in other markets, or by other firms which do not innovate, are not solutions: the correct transfer price must represent society's view of how to finance risky innovation, in an oligopoly situation with indivisible R and D costs that have to spread over a number of products, with strong elements of market power, and complicated socio-political considerations at stake. This aspect of the problem of investigating transfer pricing—which has ramifications much wider than simply collecting a fair share of taxes—has been unduly neglected in the literature (by myself, among others). Yet it is of crucial significance as far as the 'area of danger' in intra-firm trade is concerned. No solutions are proposed here, since there cannot be a correct theoretical or practical solution for high technology, high risk, innovative industries. What is proposed is that, in such cases, the issue of transfer pricing be taken out of the restrictive and misleading context of tax realization and placed in the broader, more complex, but more relevant context of paying for innovation in the framework of TNC dominated international oligopoly. The emphasis should then be on negotiation and detailed evaluation (and, inevitably, some trial and error) rather than on detecting tax evasion. Negotiation and evaluation must, in turn, be based on information, but the information by itself cannot yield a set of objectively correct transfer prices.

Negotiation on the right rate of return on risky innovation is not a task we may recommend lightly to any administration, certainly not to the overburdened officials in developing countries. Even the sophisticated and experienced office of the U.S. International Revenue Service which deals with the TNC resorts to a large number of arbitrary decisions or 'horse-trading' in cases where difficult problems of the sort mentioned above arise. Developing countries may do well in the longer run to pool their resources in order to negotiate from a stronger and better informed base, but, as matters stand, they are faced with an immensely difficult task which needs to be continuously faced. Leaving transfer prices to be decided freely by TNC is one easy way out, but it probably gives too much away. Assigning transfer prices with reference to alternative prices is another easy way out, and may indeed have no effect on innovation if a small number of developing countries practice it. As prices charged in different countries get increasingly linked to each other, especially in pharmaceuticals, this is not an ideal long-term solution. But no realistic long-term solution seems to be in sight at this time.
Concluding Remarks

In this paper we have discussed some of the important problems which arise in the investigation of transfer pricing in commodity trade from the point of view of less developed host countries. The precise relationships between the structure and nature of particular industries, the growth of intra-firm trade, the sorts of commodities that enter into such trade, and the scope for price manipulation are matters which still require much more study, from the viewpoint of policy formulation as well as because of their intrinsic theoretical interest. We have tried to make a distinction between cases where the problems raised by transfer pricing may be satisfactorily tackled with more information and those where satisfactory solutions are very difficult. The main difference between these cases lies in the nature of technological innovation: in industries with rapid and risky innovation, oligopolistic structures and considerable market power in final markets, the determination of 'correct' transfer prices is almost impossible, since it cannot be undertaken independently of the determination of the right rate of return to innovation as such. In less innovative industries, where the goods that enter into intra-firm trade are fairly standardized, it is easier to come to grips with the problem, although many developing countries have practical difficulties in collecting and assessing the information necessary to do so. The most hopeful line of action seems to lie in cooperation between countries, for problems which can be solved this could be the cheapest and most efficient solution.

3.7 Commodity Markets and Commodity Agreements

Pure competition is defined to be the situation in which no single participant has the capacity of affecting market prices more than infinitesimally (i.e., no single participant has market power). From the point of view of individual entities in the market place, prices seem to be given (not fixed over time) parameters, independent of their own behavior. Pure competition generally is considered an interesting paradigm for reasons summarized below, but not of very general applicability in the real world. However, most of the initial producers of agricultural commodities which enter into international commodity markets (and agricultural products account for about 85 percent of total non-petroleum commodity exports from developing countries) and most of the ultimate consumers of both the agricultural and non-agricultural internationally traded commodities sell and purchase these goods (generally in processed form), respectively, under conditions approximating pure competition. At both ends of the market chain for the relevant international commodities (but not in the middle, where marketing boards, other government agencies, and large companies dominate), therefore, the purely competitive model may have substantial applicability.

What are the advantages of pure competition? Some answers are in the area of political economy and, thus, derivative of particular value systems. Under pure competition, economic problems (e.g., what is produced, how it is produced, and for whom it is produced) are solved in an impersonal manner, independently of personal ties or characteristics such as race or national origin. The atomistic structure of buyers and sellers required for competition also decentralizes and disperses power. Moreover, if the conditions necessary for pure competition to exist do in fact prevail, freedom of entry into various industries and individual mobility both will be high.

Most economists focus on answers related to economic efficiency. In a world with the correct initial distribution of assets for a given social welfare function, with easy entry (e.g., due to a lack of legal restrictions and limited increasing returns to scale relative to the size of industries), with no externalities, with no uncertainty, and with pure competition everywhere else, pure competition in international commodity markets results in maximization of that social welfare function.

This is a strong result. But these are very strong conditions, obviously not even approximately satisfied in the real world. If the first one is dropped, the given social welfare function no longer is maximized, although economic efficiency remains in the Paretian sense. No shift in resources, etc., exists which would improve the welfare of any one individual without reducing the welfare of at least one other individual. Attainment of this state seems desirable, ceteris paribus, and its virtues are emphasized (perhaps overemphasized) by many economists. But without the correct distribution of assets it does not maximize a given social welfare function. Within a static framework, much of the conflict between the developing and developed countries may arise at this point. Even if all the other conditions given above are satisfied so that economic efficiency is attained, the initial distribution of assets is seen by many to be so inequitable that the world is far away from a welfare maximization. Efficiency concerns may be unimportant in light of the distribution question.

But even economic efficiency goes by the wayside once the real-world applicability of the other necessary conditions is considered. Entry is not easy in many cases due to legal, natural, and technological monopolies or due to increasing returns to scale that are large relative to the industry. The resulting market power and behavior approximating profit maximization leads to economic efficiency by driving a wedge between marginal cost and the price which the consumer uses in his or her utility maximization decision. Externalities in consumption (e.g., keeping up with the Joneses, the international demonstration effect) and in production (e.g., pollution) abound. This in itself is sufficient to preclude economic efficiency as a necessary outcome of a market system. So is the pervasive existence of uncertainty and risk aversion, with the result that on both sides of the market expected profit maximization does not necessarily lead to utility maximization.

Finally, pure competition clearly does not prevail everywhere outside of international commodity markets; therefore, the theory of the "second best" suggests that for purposes of economic efficiency, it may be preferable that it not prevail in international commodity markets either. A simple example may be useful. Assume that there are no externalities, no uncertainty, fixed factor supplies, no intermediate factors of production, and an unalterable monopoly in the rest of the world. Under these very special conditions, attempting to maintain perfect competition in international commodity markets (a "first best" policy in terms of economic efficiency) does not lead to economic efficiency, since the ratios of prices which consumers face are not equal to the ratios of marginal costs from the two sectors. However, the "second best" policy of developing the same degree of monopoly power in international markets as in the rest of the world does lead to equality in these two ratios and to economic efficiency. If specific distortions exist in the real world, second best solutions can be devised to obtain efficient outcomes. The basic problem is that devising such solutions depends upon much more knowledge than policymakers normally have.

Where does this discussion leave us? Unhindered markets may be relatively efficient devices for processing a great deal of information to signal shortages or surpluses through price or inventory changes. However, one has to be quite careful in regard to their normative implications. They lead to maximization of a given social welfare function only with the correct distribution of assets and the satisfaction of all of the other conditions discussed above. The "theory of second best" at worst implies that, in the real world, policies directed at economic efficiency should be abandoned. At best, it suggests the advocacy of "third best" very general policies which have a reasonable probability of leading to greater economic efficiency, but which do not guarantee a step in that direction when applied in a specific case.

Economic theory leads us to this highly qualified view of the normative properties of unhindered market solutions. It is a much weaker view than is suggested by many who oppose international commodity agreements on the basis that they would lessen the gains from free market operations. One can understand why economists from developing countries might question whether the position of the strongest advocate of unhindered free market operations is based on a lack of understanding of underlying economic theory or on a disguised defense of vested interests in the status quo.

Implications of Price Stabilization Attempts for Variability and Level of Revenue

The UNCTAD proposal and resolutions recognize that stabilization of export revenues probably is of much more interest to the developing countries than is stabilization of prices. In principle, of course, a buffer stock might buy and sell with the intent of stabilizing revenues. It could attempt to act so that the total market demand curve facing producers approached a unit-elastic curve with constant revenue implications. Such an operation would be much more difficult than price stabilization, however, for several reasons. Day-to-day operations would be harder because of the greater lags in the availability of quantity than price data. If such an arrangement were successful, strong inducements would exist for supply reduction since the same revenues could be earned with lower sales, which would release factors of production for other uses (although such an outcome probably could occur only if supply were organized in other than a purely competitive manner). The concurrence of importing nations with such a revenue-stabilizing scheme seems unlikely.

For such reasons, UNCTAD advocates price stabilization instead of revenue stabilization. But this strategy raises the question: What are the implications of price-stabilization attempts for revenues? Johnson states that "elementary economic analysis" suggests that the UNCTAD proposal is dubious on these grounds. He claims that under reasonable assumptions concerning the responsiveness of supply and demand to prices, there is a trade-off between the impact of price-stabilization programs on the level and on the stability of revenues. That is, increased revenues result only if there is greater instability in revenues, and vice versa. However, his argument is based on certain critical assumptions—such as high price responsiveness for suppliers and users or shifts in the underlying response relations—that do not seem justified by existing empirical evidence. Careful
consideration of the argument reveals that contrary to Johnson's "elementary economic analysis," conclusive statements may not be made on the basis of theory alone. Empirical estimates are required.

Many readers may not wish to consider further details of this issue. They may proceed to the next question of who gains from stabilization. For those who desire further clarification, a simple geometric analysis is presented in the next nine paragraphs.

Johnson's argument is illustrated in figure 3-15 below, in which the basic average supply and demand curves for a purely competitive international commodity market are given by solid straight lines, equally likely shifts in these curves are indicated by dashed lines, and $P_0$ is the average price and the price at which a buffer stock scheme would stabilize prices.

Consider first the case of instability due to shifts only in the demand curve (figure 3-15A). Without price stabilization, average producers' revenues are $(P_0Q_1 + P_0Q_2)/2$. With stabilization, they are $P_0Q_0$. Therefore, price stabilization in this case reduces the instability of producers' revenues, but also reduces their average level.

Consider next the case of instability due to supply shifts alone. Figures 3-15B-E provide illustrations of four cases with, respectively, elastic supply and demand, inelastic supply and demand, inelastic supply and elastic demand, and elastic supply and inelastic demand. Without price stabilization, producers' average revenues are $(P_2Q_1 + P_2Q_3)/2$. With stabilization, they are $P_0(Q_1 + Q_2)/2$ in figure 3-15B and $P_0(Q_2 + Q_1)/2$ in figures 3-15C-E. In each case, price stabilization increases producers' revenues.

In the case of elastic supply and demand, which Johnson considers to be most normal, however, price stabilization increases the instability of revenues (figure 3-15B). The instability of revenues also increases in two of the other cases, but not if both curves are sufficiently inelastic (figure 3-15C).

Of course, one can also consider mixed cases in which both demand and supply curves shift. The net result depends upon the size of the two shifts and other particulars of the situation. The trade-off between level and instability of revenues nevertheless seems to persist in a number of cases.

Does it follow from Johnson's conclusion that the UNCTAD advocacy of price stabilization lumps together two different economic problems (i.e., instability of demand and instability of supply) which require quite different solutions? The answer is that it depends. Suppose, for example, that the producing nations are quite risk averse. Price stabilization reduces revenue instability if demand shifts are dominant relative to supply shifts (figure 3-15A). It does the same if supply shifts are dominant and the relevant elasticities are quite low (figure 3-15C). If producers are sufficiently risk averse and if the relevant elasticities are sufficiently low, therefore, the producers are better off with price stabilization than without it no matter which curve shifts most. This example suggests that in important ways the question is an empirical one, depending on the size of the market elasticities.

The UNCTAD resolution also may make sense if shifts in one or the other of the market curves dominate. If shifts in the demand curve dominate and the producers are risk averse they may be better off even with lower average revenues, as is noted above. If shifts in the supply curve dominate, the curves are sufficiently elastic, and the producers are not very risk averse, they may be better off with the larger average revenues and greater revenue instability. If either of these cases prevails, Johnson is correct that two problems are being lumped together, but the occurrence of large shifts in one curve is sufficiently rare so that it can be ignored.

Yet another possibility is that destabilizing speculation causes large price fluctuations which lower the long-run demand curve by inducing substitution by risk-averse manufacturers of synthetics and other goods for the commodities of concern. Producers therefore might rationally prefer price stabilization in order to limit the downward long-run shift in the market demand curve, even if the short-term result is lower immediate revenues or greater instability in revenues.

These possibilities all emphasize that in important respects the matter is an empirical question. "Elementary economic analysis" is not enough. Without empirical knowledge concerning long-run movements, the shapes of the curves, risk aversion, the demand and supply elasticities of price responsiveness, the causes of shifts, whether the movements in supply and demand curves are additive or multiplicative, etc., we cannot state with assurance what the impacts of stabilization are.

**Who Gains from Price Stabilization?**

This question obviously is related to the previous discussion. Answers to it generally ignore risk, distributional aspects, and general-equilibrium effects and focus on producers' and consumers' surpluses and the financial gains from operating a buffer stock.

The most frequent situation considered is one in which instability is due to shifts in a completely inelastic supply curve. Figure 3-16 illustrates the simplest case in which the price is stabilized by buffer stock operations with no storage costs at $P_0$.

When the supply curve shifts out to $Q_1$, buffer stocks are accumulated. The consumers' benefit of paying $P_0$ instead of the price $P_1$, which would have prevailed without a buffer stock is negative ($-A$, $-B$, $-C$). The producers' benefit is the gain due
Figure 3-15. Impact on Revenues of Shifts in Demand and Supply Curves with and without Price Stabilization at $P_0$ by a Buffer Stock

A. Shift in Demand Only (elastic curves)

B. Shift in Supply Only (elastic curves)

C. Shift in Supply Only (inelastic curves)

D. Shift in Supply Only (elastic demand, inelastic supply)

E. Shift in Supply Only (inelastic demand, elastic supply)

to the higher prices ($A + B + C + D$). The cost to the buffer stock is $-C - D - E$. The total benefit (summing these three components) is $-C + E$.

When the supply curve shifts into $Q_1$, stocks are sold at $P_0$. This precludes the price from rising to $P_2$, as it otherwise would. The benefit to the consumers is $F + G$ due to the lower price and the larger quantity. The benefit to the producers is $-F$ since they receive a lower price than they would without the buffer stock. The financial inflow to the buffer stock is $B + H$. The total benefit is the sum of these three components, $B + G + H$.

If shifts in the supply curve to $Q_1$ and $Q_2$ are equally likely and if the sequencing over time is ignored (or a zero discount rate is utilized), the total benefits to each of the three groups is the sum of those obtained from buffer stock operation with supply at $Q_1$ and at $Q_2$. For consumers, the sum is $F + G - A - B - C$. For producers, the sum is $A + B + C + D - F$. For the buffer stock, the sum is $B + H - C - D - E$. The total benefit is represented by $B + G + H - C - E$. Under these assumptions, the sum for the buffer stock is zero and the overall sum is positive. However, whether consumers or producers benefit depends on the exact shape of the curves. The issue basically is an empirical one.

The provocative analysis of Newbery reinforces such an emphasis. He also explores the impact of shifts in the supply curve. He concludes that consumers’ surplus, producers’ surplus, average price, and
average supply each can increase, decrease, or remain the same under stabilization—depending on the form of the demand function (i.e., linear versus constant elasticity), the size of the price elasticity of demand, and whether the disturbance term in the supply relation is additive or multiplicative. His analysis sheds considerable doubt on the derivation of strong conclusions from too simple models. Instead it calls for empirical analysis.

But even Newbery’s extended analysis is not the whole story. As he explicitly recognizes, he does not consider the possibilities of risk aversion, distributional aspects, and general equilibrium effects. These considerations may be quite important.

But an even more important factor may be in regard to the inflation-output trade-off. What reduction in national output would have to be accepted to offset the inflationary consequences in the consuming nations of upward fluctuations in international commodity market prices together with an oligopolistic “ratchet” effect in the pricing mechanism? Existing estimates suggest that the answer may be sufficiently large to make the changes in consumers’ and producers’ surpluses second-order considerations.

**Conditions under Which Collusive Action Can Raise Market Prices**

The basic motivation behind the UNCTAD proposal and resolution may have little to do with stabilization per se. Instead, the major concern may be to raise the real resources of the developing countries that export the affected commodities.

Under certain conditions discussed above—e.g., dominant supply shifts, subject to the qualifications necessary because of Newbery’s analysis—stabilization itself may lead to increased revenues for the exporters. The content of the UNCTAD documents, however, suggests that the concern goes further than this to a desire to raise market prices (or prevent real market prices from falling) to levels above those which otherwise would prevail. If market demand curves are inelastic, successful price raising will be rewarded by greater revenues.

The question then arises, why try to form commodity agreements that include consumer representation? Why not follow the lead of OPEC? This line of inquiry leads to the further question: Under what conditions are producers likely to be able to collude by themselves to raise prices?

This leads us into a much less rigorous area of economic theory: the formation and behavior of oligopolies (i.e., a market in which there are sellers with large enough market shares that each one can recognize that his sales affect the market price). Behrman provides an extensive review of this area and its implications for modeling the structure of international commodity markets. Here I limit the presentation to a checklist of conditions which facilitate oligopolistic coordination of pricing and output decisions: (1) the perception that joint action will lead to greater returns for producers, (2) common output preferences due to similar cost structures and market shares, (3) cheap and rapid communication, (4) high concentration, (5) a small or no competitive fringe, (6) repetitive small transactions, (7) homogeneity and simplicity of products, (8) the willingness to utilize inventory and order backlogs as buffers instead of overly sensitive price adjustments, (9) limited or no substitution for the product, and (10) high barriers to entry (e.g., restricted technological knowledge, restricted control over exhaustible resources, returns to scale at a high level of production relative to market size). I also note that the dynamics of substitution and of limit pricing (i.e., pricing by a collusive group of oligopolists or by a monopolist to discourage too-rapid entry) generally leads to relatively few cases in which prices can be long sustained significantly above marginal costs.

Before turning to the empirical evidence, however, it is useful to consider briefly two more topics related to the existence of market power in international primary commodity markets.

First, whatever reorganization of the market structure is proposed by UNCTAD or by anyone else, it is obvious that the starting point is not purely competitive international commodity markets. To the contrary, there are in existence a number of private
and national-government organizations that currently can exert market power in international commodity markets. Thus Johnson's elementary economic analysis may be misleading because it assumes the wrong market structure to start with. Indeed, from this point of view, price stabilization schemes may be a device to increase oligopolistic price discipline (although producer cartels might be as effective as international commodity agreements in attaining this end).

Second, considerable attention recently has been paid to the case of exhaustible resources, a category into which most of the important mineral primary commodities usually are placed. There seems to be a fairly widespread notion that such pricing will be far from competitive levels. Stiglitz recently compared the rate of exploitation and the path of prices of an exhaustible resource in competitive markets with that for a profit-maximizing monopolist. It is perhaps surprising to some that his analysis suggests that there is very little scope for the monopolist to exercise monopoly power. Under the assumptions of constant elasticity of demand schedules and zero marginal costs of extraction, monopoly prices and rates of extraction are identical to those of the competitive equilibrium! This result contrasts with the higher prices and lower production rates for monopoly than

Externalities

By "externalities" the economist means a side effect or spillover in resource use that falls outside the market mechanism and hence goes unpriced. An externality results when the activity of a business firm, a household, or a public agency affects the activities of other firms, households, or public agencies that are exerted otherwise than through the relationships between buyer and seller in a market. There are then benefits or costs for third parties that are not counted by the market-price system. The externality can be positive—a spillover that lowers costs elsewhere, that is, an external economy. Or it may be negative—a spillover that raises costs elsewhere, that is, an external diseconomy or detriment.

Some common examples of external diseconomies are the damages inflicted on others through various forms of pollution, congestion in road use, use of water from wells that all tap a common source of water, or side effects from irrigation schemes on health or fisheries.

The social cost of a firm's activity equals the private cost to the firm plus any external diseconomy caused by the firm's activity. The social value equals the market price less any external detriment imposed on a nonuser. But since the firm does not have to pay for the external diseconomy, the market price (equal only to private cost) no longer indicates the social cost of the activity or the social value. If the cost of the external detriment cannot be imposed on the firm—that is, "internalized"—the firm will underestimate the true cost of its activity, and it will use too many resources in this activity. The social value of the good provided by this activity—namely, its market price minus the value of
the damage inflicted on others by producing the goods—will be below its market price, and may even be negative. Even if perfect competition existed, but the social cost exceeded private cost, there would be a misallocation of resources. (Price would equal private marginal cost but not social marginal cost.) To remove the overuse of resources in an activity that creates external diseconomies, the government might tax the activity or regulate it by licensing or establishing specific standards. It might also be possible to internalize the externalities by centralizing decisionmaking within a group of units that were formerly separate decisionmaking units. These corrective remedies will reduce output until the social value of the good is raised sufficiently to become equal to its marginal cost of production.

On the benefit side, one firm's training of labor that becomes available to another firm, or the demonstration effects of a new product or process, or the introduction of a transport network that becomes freely available to others are all examples of external economies or nonuser benefits that cannot be appropriated by the firm that creates the side benefit. The firm therefore underestimates the benefit of its activity; social cost now equals private cost minus any external economy; or social value of the activity exceeds its market price by the additional benefit conferred elsewhere. If external economies are created, the firm should be encouraged to produce more than it would by only private marginal cost considerations. Remedial policy would therefore constitute a subsidy to the firm to lower private marginal cost to social marginal cost.

The choice of policy instruments to deal with externalities deserves more attention. It is all very easy to say in a formal sense that resources should be devoted to an activity until the expected marginal social benefit of that activity equals the marginal social benefit in alternative uses. But we are still left with the operational question of just what remedial policies should be instituted to correct market failure.

Selection 3.8 considers the main alternative policy measures for dealing with external diseconomies, particularly regulation through direct controls and financial incentives through pricing. Selection 3.9 then considers the evaluation of external economies for developing countries, especially in the case of industry.

3.8 The Problem of Social Cost

This paper is concerned with those actions of business firms which have harmful effects on others. The standard example is that of a factory, the smoke from which has harmful effects on those occupying neighbouring properties. The economic analysis of such a situation has usually proceeded in terms of a divergence between the private and social product of the factory, in which economists have largely followed the treatment of Pigou in The Economics of Welfare. The conclusions to which this kind of analysis seems to have led most economists is that it would be desirable to make the owner of the factory liable for the damage caused to those injured by the smoke, or alternatively, to place a tax on the factory owner varying with the amount of smoke produced and equivalent in money terms to the damage it would cause, or finally, to exclude the factory from residential districts (and presumably from other areas in which the emission of smoke would have harmful effects on others). It is my contention that the suggested courses of action are inappropriate, in that they lead to results which are not necessarily, or even usually, desirable.

The Reciprocal Nature of the Problem

The traditional approach has tended to obscure the nature of the choice that has to be made. The question is commonly thought of as one in which A inflicts harm on B and what has to be decided is: how should we restrain A? But this is wrong. We are
dealing with a problem of a reciprocal nature. To avoid the harm to B would inflict harm on A. The real question that has to be decided is: should A be allowed to harm B or should B be allowed to harm A? The problem is to avoid the more serious harm. I instanced in my previous article the case of a confectioner the noise and vibrations from whose machinery disturbed a doctor in his work. To avoid harming the doctor would inflict harm on the confectioner. The problem posed by this case was essentially whether it was worth while, as a result of restricting the methods of production which could be used by the confectioner, to secure more doctoring at the cost of a reduced supply of confectionery products. Another example is afforded by the problem of straying cattle which destroy crops on neighbouring land. If it is inevitable that some cattle will stray, an increase in the supply of meat can only be obtained at the expense of a decrease in the supply of crops. The nature of the choice is clear: meat or crops. What answer should be given is, of course, not clear unless we know the value of what is obtained as well as the value of what is sacrificed to obtain it. To give another example, Professor George J. Stigler instances the contamination of a stream. If we assume that the harmful effect of the pollution is that it kills the fish, the question to be decided is: is the value of the fish lost greater or less than the value of the product which the contamination of the stream makes possible. It goes almost without saying that this problem has to be looked at in total and at the margin.

The Pricing System with Liability for Damage

I propose to start my analysis by examining a case in which most economists would presumably agree that the problem would be solved in a completely satisfactory manner: when the damaging business has to pay for all damage caused and the pricing system works smoothly (strictly this means that the operation of a pricing system is without cost).

A good example of the problem under discussion is afforded by the case of straying cattle which destroy crops growing on neighbouring land. Let us suppose that a farmer and a cattle-raiser are operating on neighbouring properties. Let us further suppose that, without any fencing between the properties, an increase in the size of the cattle-raiser's herd increases the total damage to the farmer's crops. What happens to the marginal damage as the size of the herd increases is another matter. This depends on whether the cattle tend to follow one another or to roam side by side, on whether they tend to be more or less restless as the size of the herd increases, and on other similar factors. For my immediate purpose, it is immaterial what assumption is made about marginal damage as the size of the herd increases.

To simplify the argument, I propose to use an arithmetical example. I shall assume that the annual cost of fencing the farmer's property is $9 and that the price of the crop is $1 per ton. Also, I assume that the relation between the number of cattle in the herd and the annual crop loss is as follows:

<table>
<thead>
<tr>
<th>Number in herd (steers)</th>
<th>Annual crop loss (tons)</th>
<th>Crop loss per additional steer (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

Given that the cattle-raiser is liable for the damage caused, the additional annual cost imposed on the cattle-raiser if he increased his herd from, say, 2 to 3 steers is $3 and in deciding on the size of the herd, he will take this into account along with his other costs. That is, he will not increase the size of the herd unless the value of the additional meat produced (assuming that the cattle-raiser slaughters the cattle), is greater than the additional costs that this will entail, including the value of the additional crops destroyed. Of course, if, by the employment of dogs, herdsman, aeroplanes, mobile radio and other means, the amount of damage can be reduced, these means will be adopted when their cost is less than the value of the crop which they prevent being lost. Given that the annual cost of fencing is $9, the cattle-raiser who wished to have a herd with 4 steers or more would pay for fencing to be erected and maintained, assuming that other means of attaining the same end would not do so more cheaply. When the fence is erected, the marginal cost due to the liability for damage becomes zero, except to the extent that an increase in the size of the herd necessitates a stronger and therefore more expensive fence because more steers are liable to lean against it at the same time. But, of course, it may be cheaper for the cattle-raiser not to fence and to pay for the damaged crops, as in my arithmetical example, with 3 or fewer steers.

It might be thought that the fact that the cattle-raiser would pay for all crops damaged would lead the farmer to increase his planting if a cattle-raiser came to occupy the neighbouring property. But this is not so. If the crop was previously sold in conditions of perfect competition, marginal cost was equal to price for the amount of planting undertaken and any expansion would have reduced the profits of the farmer. In the new situation, the existence of crop damage would mean that the farmer would sell less on the open market but his receipts for a given production would remain the same, since the cattle-raiser would pay the market price for any crop damaged. Of course,
if cattle-raising commonly involved the destruction of crops, the coming into existence of a cattle-raising industry might raise the price of the crops involved and farmers would then extend their planting. But I wish to confine my attention to the individual farmer.

I have said that the occupation of a neighbouring property by a cattle-raiser would not cause the amount of production, or perhaps more exactly the amount of planting, by the farmer to increase. In fact, if the cattle-raising has any effect, it will be to decrease the amount of planting. The reason for this is that, for any given tract of land, if the value of the crop damaged is so great that the receipts from the sale of the undamaged crop are less than the total costs of cultivating that tract of land, it will be profitable for the farmer and the cattle-raiser to make a bargain whereby that tract of land is left uncultivated. This can be made clear by means of an arithmetical example. Assume initially that the value of the crop obtained from cultivating a given tract of land is $12 and that the cost incurred in cultivating this tract of land is $10, the net gain from cultivating the land being $2. I assume for purposes of simplicity that the farmer owns the land. Now assume that the cattle-raiser starts operations on the neighbouring property and that the value of the crops damaged is $1. In this case $11 is obtained by the farmer from sale on the market and $1 is obtained from the cattle-raiser for damage suffered and the net gain remains $2.

Now suppose that the cattle-raiser finds it profitable to increase the size of his herd, even though the amount of damage rises to $3; which means that the value of the additional meat production is greater than the additional costs, including the additional $2 payment for damage. But the total payment for damage is now $3. The net gain to the farmer from cultivating the land is still $2. The cattle-raiser would be better off if the farmer would agree not to cultivate his land for any payment less than $3. The farmer would be agreeable to not cultivating the land for any payment greater than $2. There is clearly room for a mutually satisfactory bargain which would lead to the abandonment of cultivation. But the same argument applies not only to the whole tract cultivated by the farmer but also to any cases in which the damage that the cattle would cause, and for which the cattle-raiser would be willing to pay, exceeds the amount which the farmer would pay for use of the land. In conditions of perfect competition, the amount which the farmer would pay for the use of the land is equal to the difference between the value of the total production when the factors are employed on this land and the value of the additional product yielded in their next best use (which would be what the farmer would have to pay for the factors). If damage exceeds the amount the farmer would pay for the use of the land, the value of the additional product of the factors employed elsewhere would exceed the value of the total product in this use after damage is taken into account. It follows that it would be desirable to abandon cultivation of the land and to release the factors employed for production elsewhere. A procedure which merely provided for payment for damage to the crop caused by the cattle but which did not allow for the possibility of cultivation being discontinued would result in too small an employment of factors of production in cattle-raising and too large an employment of factors in cultivation of the crop. But given the possibility of market transactions, a situation in which damage to crops exceeded the rent of the land would not endure. Whether the cattle-raiser pays the farmer to leave the land uncultivated or himself rents the land by paying the land-owner an amount slightly greater than the farmer would pay (if the farmer was himself renting the land), the final result would be the same and would maximize the value of production. Even when the farmer is induced to plant crops which it would not be profitable to cultivate for sale on the market, this will be a purely short-term phenomenon and may be expected to lead to an agreement under which the planting will cease. The cattle-raiser will remain in that location and the marginal cost of meat production will be the same as before, thus having no long-run effect on the allocation of resources.

The Pricing System with No Liability for Damage

I now turn to the case in which, although the pricing system is assumed to work smoothly (that is, cost-

---

14. The argument in the text has proceeded on the assumption that the alternative to cultivation of the crop is abandonment of cultivation altogether. But this need not be so. There may be crops which are less liable to damage by cattle but which would not be as profitable as the crop grown in the absence of damage. Thus, if the cultivation of a new crop would yield a return to the farmer of $1 instead of $2, and the size of the herd which would cause $3 damage with the old crop would cause $1 damage with the new crop, it would be profitable to the cattle-raiser to pay any sum less than $2 to induce the farmer to change his crop (since this would reduce damage liability from $3 to $1) and it would be profitable for the farmer to do so if the amount received was more than $1 (the reduction in his return caused by switching crops). In fact, there would be room for a mutually satisfactory bargain in all cases in which a change of crop would reduce the amount of damage by more than it reduces the value of the crop (excluding damage)—in all cases, that is, in which a change in the crop cultivated would lead to an increase in the value of production.
lessly), the damaging business is not liable for any of
the damage which it causes. This business does not
have to make a payment to those damaged by its
actions. I propose to show that the allocation of
resources will be the same in this case as it was when
the damaging business was liable for damage caused.
As I showed in the previous case that the allocation
of resources was optimal, it will not be necessary to
repeat this part of the argument.

I return to the case of the farmer and the cattle­
raiser. The farmer would suffer increased damage to
his crop as the size of the herd increased. Suppos:
that the size of the cattle-raiser’s herd is 3 steers (and
that this is the size of the herd that would be main­
tained if crop damage was not taken into account).
Then the farmer would be willing to pay up to $3 if
the cattle-raiser would reduce his herd to 2 steers,
up to $5 if the herd were reduced to 1 steer and
would pay up to $6 if cattle-raising was abandoned.
The cattle-raiser would therefore receive $3 from
the farmer if he kept 2 steers instead of 3. This $3 forgone
is therefore part of the cost incurred in keeping the
third steer. Whether the $3 is a payment which the
cattle-raiser has to make if he adds the third steer to
his herd (which it would be if the cattle-raiser were
liable to the farmer for damage caused to the crop)
or whether it is a sum of money which he would have
received if he did not keep a third steer (which
it would be if the cattle-raiser were not liable to the
farmer for damage caused to the crop) does not affect
the final result. In both cases $3 is part of the cost
of adding a third steer, to be included along with the
other costs. If the increase in the value of production
in cattle-raising through increasing the size of the
herd from 2 to 3 is greater than the additional costs
that have to be incurred (including the $3 damage
to crops), the size of the herd will be increased.
Otherwise, it will not. The size of the herd will be
the same whether the cattle-raiser is liable for damage
caused to the crop or not.

It may be argued that the assumed starting point—a
herd of 3 steers—was arbitrary. And this is true.
But the farmer would not wish to pay to avoid crop
damage which the cattle-raiser would not be able to
cause. For example, the maximum annual payment
which the farmer could be induced to pay could not
exceed $9, the annual cost of fencing. And the farmer
would only be willing to pay this sum if it did not
reduce his earnings to a level that would cause him
to abandon cultivation of this particular tract of land.
Furthermore, the farmer would only be willing to pay this
amount if he believed that, in the absence of
any payment by him, the size of the herd maintained
by the cattle-raiser would be 4 or more steers. Let us
assume that this is the case. Then the farmer would
be willing to pay up to $3 if the cattle-raiser would
reduce his herd to 3 steers, up to $6 if the herd were
reduced to 2 steers, up to $8 if one steer only were
kept, and up to $9 if cattle-raising were abandoned.
It will be noticed that the change in the starting
point has not altered the amount which would accrue
to the cattle-raiser if he reduced the size of his herd
by any given amount. It is still true that the cattle­
raiser could receive an additional $3 from the farmer
if he agreed to reduce his herd from 3 steers to 2 and
that the $3 represents the value of the crop that
would be destroyed by adding the third steer to the
herd. Although a different belief on the part of the
farmer (whether justified or not) about the size of
the herd that the cattle-raiser would maintain in the
absence of payments from him may affect the total
payment he can be induced to pay, it is not true that
this different belief would have any effect on the size
of the herd that the cattle-raiser will actually keep.
This will be the same as it would be if the cattle­
raiser had to pay for damage caused by his cattle,
since a receipt forgone of a given amount is the
equivalent of a payment of the same amount.

It might be thought that it would pay the cattle­
raiser to increase his herd above the size that he
would wish to maintain once a bargain had been made,
in order to induce the farmer to make a larger
total payment. And this may be true. It is similar in
nature to the action of the farmer (when the cattle­
raiser was liable for damage) in cultivating land on
which, as a result of an agreement with the cattle­
raiser, planting would subsequently be abandoned
(including land which would not be cultivated at
all in the absence of cattle-raising). But such
manoeuvres are preliminaries to an agreement and
do not affect the long-run equilibrium position, which
is the same whether or not the cattle-raiser is held
responsible for the crop damage brought about by his
cattle.

It is necessary to know whether the damaging busi­
ness is liable or not for damage caused since without
the establishment of this initial delimitation of rights
there can be no market transactions to transfer and
recombine them. But the ultimate result (which
maximizes the value of production) is independent
of the legal position if the pricing system is assumed
to work without cost.

A Change of Approach

It is my belief that the failure of economics to reach
correct conclusions about the treatment of harmful
effects cannot be ascribed simply to a few slips in
analysis. It stems from basic defects in the current
approach to problems of welfare economics. What is
needed is a change of approach.

Analysis in terms of divergences between private
and social products concentrates attention on par-
ticular deficiencies in the system and tends to nourish the belief that any measure which will remove the deficiency is necessarily desirable. It diverts attention from those other changes in the system which are inevitably associated with the corrective measure, changes which may well produce more harm than the original deficiency. In the preceding sections of this article, we have seen many examples of this. But it is not necessary to approach the problem in this way. Economists who study problems of the firm habitually use an opportunity cost approach and compare the receipts obtained from a given combination of factors with alternative business arrangements. It would seem desirable to use a similar approach when dealing with questions of economic policy and to compare the total product yielded by alternative social arrangements. In this article, the analysis has been confined, as is usual in this part of economics, to comparisons of the value of production, as measured by the market. But it is, of course, desirable that the choice between different social arrangements for the solution of economic problems should be carried out in broader terms than this and that the total effect of these arrangements in all spheres of life should be taken into account. As Frank H. Knight has so often emphasized, problems of welfare economics must ultimately dissolve into a study of aesthetics and morals.

A second feature of the usual treatment of the problems discussed in this article is that the analysis proceeds in terms of a comparison between a state of laissez faire and some kind of ideal world. This approach inevitably leads to a looseness of thought since the nature of the alternatives being compared is never clear. In a state of laissez faire, is there a monetary, a legal or a political system and if so, what are they? In an ideal world, would there be a monetary, a legal or a political system and if so, what would they be? The answers to all these questions are shrouded in mystery and every man is free to draw whatever conclusions he likes. Actually very little analysis is required to show that an ideal world is better than a state of laissez faire, unless the definitions of a state of laissez faire and an ideal world happen to be the same. But the whole discussion is largely irrelevant for questions of economic policy since whatever we may have in mind as our ideal world, it is clear that we have not yet discovered how to get to it from where we are. A better approach would seem to be to start our analysis with a situation approximating that which actually exists, to examine the effects of a proposed policy change and to attempt to decide whether the new situation would be, in total, better or worse than the original one. In this way, conclusions for policy would have some relevance to the actual situation.

A final reason for the failure to develop a theory adequate to handle the problem of harmful effects stems from a faulty concept of a factor of production. This is usually thought of as a physical entity which the businessman acquires and uses (an acre of land, a ton of fertiliser) instead of as a right to perform certain (physical) actions. We may speak of a person owning land and using it as a factor of production but what the land-owner in fact possesses is the right to carry out a circumscribed list of actions. The rights of a land-owner are not unlimited. It is not even always possible for him to remove the land to another place, for instance, by quarrying it. And although it may be possible for him to exclude some people from using "his" land, this may not be true of others. For example, some people may have the right to cross the land. Furthermore, it may or may not be possible to erect certain types of buildings or to grow certain crops or to use particular drainage systems on the land. This does not come about simply because of government regulation. It would be equally true under the common law. In fact it would be true under any system of law. A system in which the rights of individuals were unlimited would be one in which there were no rights to acquire.

If factors of production are thought of as rights, it becomes easier to understand that the right to do something which has a harmful effect (such as the creation of smoke, noise, and smells) is also a factor of production. Just as we may use a piece of land in such a way as to prevent someone else from crossing it, or parking his car, or building his house upon it, so we may use it in such a way as to deny him a view or quiet or unpolluted air. The cost of exercising a right (of using a factor of production) is always the loss which is suffered elsewhere in consequence of the exercise of that right—the inability to cross land, to park a car, to build a house, to enjoy a view, to have peace and quiet, or to breathe clean air.

It would clearly be desirable if the only actions performed were those in which what was gained was worth more than what was lost. But in choosing between social arrangements within the context of which individual decisions are made, we have to bear in mind that a change in the existing system which will lead to an improvement in some decisions may well lead to a worsening of others. Furthermore we have to take into account the costs involved in operating the various social arrangements (whether it be the working of a market or of a government department), as well as the costs involved in moving to a new system. In devising and choosing between social arrangements we should have regard for the total effect. This, above all, is the change in approach which I am advocating.
3.9 Externalities in Developing Countries

The concept 'external economies' has been widely used in cases where the social profitability of a project is thought to be higher than the profitability from the point of view of the enterprise: sometimes rather regardless of the exact reason for such a supposed difference. Similarly, 'external diseconomies' may be used to refer to the opposite case. But it is the external economies which have been emphasized for developing countries: especially in the case of industry, in defence against criticism that industrialization is overemphasized in developing countries. In this chapter we shall not concern ourselves with all such differences, for the good reason that our system of cost-benefit analysis is precisely supposed to allow for many of them. The question that arises is therefore 'what costs and benefits, if any, have escaped the analysis which has been propounded?'

We have proposed a system of valuing the inputs and outputs of a project according to a set of accounting prices supposed to measure social costs and benefits. The question of the previous paragraph can thus be divided into two questions:

(1) Are there inputs and outputs which we have failed to include—what we may call extraordinary inputs and outputs to distinguish them from the ordinary ones which we have certainly included?

(2) Have we misvalued the obvious inputs and outputs, because they themselves have extraordinary benefits or costs for society?

Certain external economies are sometimes attributed to industrialization in general; for example, the inculcation of disciplined working habits: these are closely related to those external economies which are required to validate what is known as the 'infant industry' argument, an argument which has long been used to justify the protection of, and special fiscal treatment for, any newly established industries or firms. In brief, the argument relies, mainly if not wholly, on the premise that the entrepreneur in a new line of activity is inadequately rewarded for the high cost of acquiring technical knowledge and skills (including the training of managers and workers), because such learning and training has benefits which he cannot keep to himself, and charge for. This kind of externality is discussed in the next section, which deals with outputs.

External Economies Related to Outputs

EXTRAORDINARY OUTPUTS

As already suggested, the first possibility is that there are outputs not ordinarily counted as such, which may consequently be overlooked. The most obvious case in industry of a beneficial extraordinary output seems to be labour-training (including skilled labour). In general, people improve their skills by being employed, in a manner which increases their value to other employers, or sometimes to themselves if they leave and set up their own business. When a man leaves a firm, the latter will thus have added some value to him. Is it in any way recompensed for this 'product'? The answer is, only to the extent that it may have got the man cheap in the first place in anticipation of the training (e.g., apprenticeship). For even if the man stays with the firm that trains him, the latter is likely to have to pay him more as a result of the training it itself has provided, in order to keep him. It should be possible to make a rough quantitative estimate of this external economy when appraising a project. The main question is whether it is worth doing. Our impression is that the present value of such training is usually likely to be small compared with other project items.

The second possibility is the cornerstone of the infant industry argument. If a firm starts a new product or new process, it may make it easier for other firms to do likewise, and the initiator's benefit may then fall short of the social benefit. Patent laws are intended to ensure that the inventor or patent-purchaser is sufficiently protected, by ensuring that external economies are not realized, for a while. But the industrial innovator relies very often more on secrecy than on patents, and on retaining a monopoly of the 'know-how,' sometimes with the aid of (probably unenforceable) service agreements. Even when there is no mystery, and little problem in acquiring know-how (and therefore no question of a leakage), the successful innovator may, especially in a developing country, stimulate others to try their hands at the same game.

In the industrialized countries, much more attention is now paid to extraordinary bad industrial outputs than to possible external benefits. Pollution has found its way from academic text-books to the popular imagination. Developing countries are nervous of this change of emphasis, and some regard it as a phenomenon only of a high level of industrialization (although it is, perhaps, now more obvious in some

towns in the developing world than in most of those of the industrialized countries).

Turning to agriculture, it seems to us that external economies at the farm level are more obvious than in industry. The classic example of an external economy—the fruit-farmers who produce blossom and so aid bee-keepers—comes from agriculture, just as the classic example of an external diseconomy—smoke—comes from industry. More important, it appears that the spread of know-how in agriculture comes from direct imitation to a greater extent than in the case of industry. This is sometimes implicitly recognized when new agricultural inputs are subsidized: while the most important output of a demonstration farm is supposed to be its external effect. Equally one can say that the externality of training in industry is recognized by the institution of training centres, institutes of technology, and management schools.

**Undervaluation of Ordinary Outputs**

It is quite common for certain outputs to be sold for less than their social value, for administrative or political reasons. This is most usual in the case of infrastructural projects. Thus roads are usually supplied free. Irrigation water is either free, or sold for less than its worth. Public electricity undertakings as often as not lose money. No one conducting a cost-benefit analysis would think of ignoring such divergences between the price charged and the social value of the output, which must then be estimated by some other means. In manufacturing and agriculture, deliberate undercharging is less common. It is also unusual for it to be administratively difficult to charge people as much as they would be willing to pay.

More subtle reasons why one might underestimate the value of some industrial outputs have been suggested. These suggestions concern the domestic output of intermediate and capital goods. In this connection, the concept of ‘forward linkages’ has received wide attention, a ‘forward linkage’ being a relationship between two firms considered from the seller’s point of view, while a ‘backward’ linkage is a relationship seen from the buyer’s point of view.

Consider a new project which uses a cost-reducing innovation. Now it is possible that in practice some of the benefit of the cost reduction will be passed on (or forward) to other industries. But there is no reason, so far, why a project analysis should go wrong: the output would be normally valued at the old price, which would properly account for the benefit. It has been suggested, not so much that the benefit of a given project may be underestimated, as that the size of the investment is likely to be too small, because increased demand arising as a result of the cost reductions passed on to other firms or industries (called, sometimes, pecuniary external economies) will not be allowed for by private industry. This argument raises the whole question of investment planning, and of whether, when information is inevitably limited, planning or the price mechanism produce better results for investment. But it does not seem to be a very strong argument in the context of external economies; for investments generally anticipate increasing demand, and that part of the increased demand which is caused indirectly by the investment is likely to be inconsiderable in relation to the total increase. In investment planning, much depends on the market conditions facing the investor, and his attitude. At all events, excessive investment in capacity often arises in industries where one might expect the opposite, according to the above argument. There seems to be nothing in this argument which should make us want to modify what has been said: that demand should be estimated as well as it can be, with or without the help of planning.

Still on the subject of ‘forward linkages,’ it has been maintained that the production of intermediates encourages others to make investments which use these inputs, and that this is an advantage. What is undeniable is that lack of an input can inhibit otherwise profitable and socially beneficial investment. This is the essential argument for providing a suitable infrastructure, which can be commonly used by manufacturing firms, in the case of non-traded inputs which can best be produced on a scale which requires their use by many customers. But there is nothing here which implies that the rules suggested for the production of non-traded inputs require any modification. Such things as transport and power should be available at prices which equal the marginal social cost or benefit of providing them. But it should not be thought that the mere provision of such things will result in any use being made of them. There are a number of monuments to this mistake in developing countries.

**External Economies Related to Inputs**

**Extraordinary Inputs**

In theory, if a project receives some beneficial inputs for which it does not pay, this should be counted as a cost, and the producer of the benefit should receive an equal recompense; and similarly if it is harmed, e.g., by another's smoke, its costs should be reduced by an appropriate amount. We are not aware of cases where such adjustments would have a very important effect on costs.

**Wrong Valuation of Inputs**

It is possible that the general development of the economy will reduce the real cost of some of the
inputs of a project. But this is a matter of proper estimation of future accounting prices and is not an 'external effect,' unless the initiation of the project itself causes changes in the real cost of its inputs. This can happen if the demand of the project for an input is either (a) sufficient to result in the establishment of a socially profitable project to produce the input, or (b) sufficient to realize economies in the production of the input, when the latter is already produced domestically. These kinds of effect, via new demand for inputs, are often referred to as economies resulting from backward linkages.

The first case could arise if there was previously no domestic demand for the input, while it could not have been produced for export with adequate social profitability because of transport costs or foreign protection. But, given a domestic demand, the input can be produced at a lower social accounting cost than the import price. Now the project under consideration might show inadequate social profitability if the input were reckoned at the c.i.f. price (there being no domestic production); but it would pass the test if, anticipating its production, one treated the input as a non-traded good (ex hypothesi its accounting cost is then lower than the c.i.f. price).

To illustrate the above paragraph, it might be socially profitable to produce refrigerators together with electric motors; but if one evaluated a refrigerator project based on imported motors, and an electric motor project without the refrigerator demand, neither separately would pass the test.

Now, supposing that the refrigerator project is under examination, the methods of evaluation proposed need no modification, provided that it is realized that domestic electric motor production will spring up as a result: given this, the proper accounting price will be put on the motors. If, say, the main project (refrigerators) is in the public sector, and the input-supplying project or projects will be in the private sector, it may be as well to ensure that it or they in fact get started by offering long-term contracts, credit, etc. If there is no question of either project being in the public sector, it can happen that neither gets started since each waits upon the other. In this case no question of project appraisal will arise, but an opportunity is missed. But that such opportunities can be missed is no proof that they often are; after all, either the refrigerator man can decide to make motors himself, or the motor man can make refrigerators. True, such 'vertical integration' requires more capital, and this might be a stumbling block: on the other hand, if each of the men can raise capital for his own part of the business, it is not inconceivable that they should amalgamate their projects, or at least make a contract.

A very similar situation arises when some of the inputs of the project are already made at home. In this case, the increased demand for them may result in economies of scale, and these economies may be external to the project. They will be external if the resulting benefit is not passed back in the form of lower prices. Does our method of project appraisal automatically take account of this kind of benefit? Probably 'yes,' but let us consider the matter further. Suppose some of the input is imported and will continue to be so for some time. In this case, it is valued at its c.i.f. price, which makes no allowance for induced economies of scale; but this is correct, for if the good is still imported, then the economies of scale could in any case have been realized by further import substitution, and are not attributable to the project. On the other hand, if the import price tends to fall as less is imported, complete import substitution may not have been a correct policy despite domestic economies of scale (in this case we would have used the marginal cost of importing as the basis for the accounting price of the input, and not the c.i.f. price). Similarly, when the input is an export commodity, economies of scale can arise as a result of the project, if the export market is not perfectly elastic. Economies of scale can also arise if the good is not traded at all. Thus the question whether our methods already take account of such economies boils down to the question of whether they will be correctly anticipated, if they arise, and so be allowed for in the future accounting prices of inputs. There is really no matter of principle involved. The more important question is whether they are likely to be very significant, and whether it is sensible to spend much time on enquiries and research which would enable one to make some sort of rough estimate.

Much has been made of the benefits to be expected to result from backward linkages. It has often been claimed that a project which was manifestly socially unprofitable at the time would become socially profitable when the heavy dependence on imported components was reduced—without any evidence being offered, as if the matter was self-evident. Again, it is sometimes argued that motor-car assembly must be a good thing for developing countries to promote, because it leads to the manufacture of the component parts. Governments may require that increasing percentages of the value of the final product result from domestic manufacture as time goes on.

Some warnings are in order. If the establishment of backward linkages is sought for its own sake, without economic appraisal, then the social profitability of a project, whose costs consist to a considerable extent of the purchases of parts, is likely to be reduced, not raised. For instance, some parts may need to be made on a much larger scale than is required for their assembly into a particular final product. This seems to be notably the case in the motor-car industry. It is for this reason that it is often referred to
as essentially an assembly industry. It is also worth
remarking that it seems to be increasingly the case,
even in industrialized countries, that imported com-
ponents are incorporated in a final 'national' product.
Even so considerable a component as the engine is
remarking that it seems to be increasingly the case,
important to add the slogan 'trade in intermediates.' Finally, it
should not be imagined, even when it seems likely
to automatic about the development of backward linkages. Sometimes it happens; but often it does not.

We shall end this section by asking whether there
is any way of deciding a priori how important the
external effects arising from backward linkages might
be. Let us make up a very simple hypothetical project.
All figures are in accounting prices. Rs50 million of
capital is spent in year 1: thereafter in every year
forever, sales are Rs15 million and current costs are
Rs10 million. Discounted at 10 percent this flow of
social profit gives Rs50 million of present value, equal
to the capital expenditure. Hence the present social
value (psv) is zero.

Now suppose that 10 percent of current costs is in
respect of purchases of a component, not subject to
economies of scale, which was previously wholly im-
ported, but which is now going to be made domes-
tically as a result of our project. The project thus
spends Rs1 million per annum on this item, at c.i.f.
prices. Now the new domestic production is unlikely
to reduce social costs by more than, say, 15 percent.
If the component could have been made much more
cheaply than this, it is reasonable to assume that it
would have been worth making for export. Let us
suppose that production of this component starts at
the same time as the project. This is a very favourable
assumption, because if this really were to happen its
production would have to be planned in conjunction
with the project—in which case any sensible project
appraisal would reckon in these external benefits,
which we are now assuming to be forgotten or at
least ignored. On these assumptions, annual current
costs are overestimated by Rs150,000 per annum.

Assume now that a further 10 percent of costs is in
respect of purchases of inputs which are not spe-
cific to the project (i.e., have other uses), where
economies of scale are consequently realized. The
increased demand is thus Rs1 million. How large a
percentage increase in output of the item or items
this represents, depends on the previous size of the
market. It would seem to be extreme to assume that
average costs might fall 20 percent as a result of such
increased demand. Nevertheless, this might happen
if the increased demand was, say, in respect of one
item whose output was thereby doubled: so this is
the assumption we shall make. But Rs200,000 of
consequent saving cannot be reckoned for every year
of the life of our project. Since the item is non-
specific, the general development of the economy
would have caused the realization of the economies
of scale anyway. The project under examination only
advances the realization. Its contribution to these
economies of scale thus falls away from the maximum
figure of, say, 20 percent after a year or two (when
the new capacity is created to meet the increased
demand caused by the project) to a negligible amount
after, say, 20 years. As a rough allowance for both
the delay in the realization of such economies, and
their decreasing importance, one might reckon that
the 20 percent mentioned is the equivalent of 10
percent for every year of the life of the project. Con-
sequently another external economy of Rs100,000 is
realized.

This adds up to a total of Rs250,000 external
economies per annum forever: which compares with
projected sales of Rs15 million, a projection which
is most likely to have a range of at least plus or minus
10 percent, i.e., Rs1.5 million. This example, de-
dsigned to show how one might make a quantitative
estimate of this particular externality, does, despite
its hypothetical nature, make one wonder whether
it is worth spending a lot of time on trying to estimate
external economies. It may be far more important to
spend the time improving the ordinary estimates of
sales and costs.

Industrial and Spatial Complexes

Both of these have been mentioned en passant. The
manufacture of refrigerators and electric motors could
be considered as a rudimentary 'complex.' Any set of
plants such that one buys most of the output of
all the others, or all but one or two plants sell most
of their output to another member or members of the
set, seems to be what is meant by this rather vague
term. People, for instance, speak of a petro-chemical
complex. Where such a set of linked projects can be
set up, it may be advisable to do a cost-benefit ap-
praisal of the whole complex. This is because the
situation already discussed in the refrigerator and
electric motor example may arise. In other words, a
set of plants may be sufficiently socially profitable;
buts, taken one by one, and without the local market
provided by other plants, no constituent plant would
be socially profitable. In extreme cases, this is ob-
vious enough, as when a product is very expensive
to transport, and can be used only as an input into
another process—e.g., some of the gases in a petro-
chemical complex. But, with such a complex, it is
also advisable to look at each plant separately when-
ever the result is not obvious—whenever, that is, an input can be purchased outside the complex, or an intermediate output sold outside it. The complex should not be regarded as technically determined. Sometimes, an input might be better imported. Sometimes, it might be profitable to produce more of some intermediate than needed by the complex, and sell it as well as the final product.

We have also briefly mentioned the economies that can result from physical proximity, independently of vertical technological linkages. This is what we mean by spatial complexes. If one is designing a spatial industrial complex from scratch, like decorating a completely bare room, then it becomes useful to know what benefits derive from proximity. What industries or plants gain by being close to each other? There appears to be very little empirical work indeed which helps to answer this question. Ordinary transport economies are obvious enough; so also are the diseconomies, e.g., locating oneself near any input may mean increasing one’s distance from the market for the final product. Other slightly less obvious economies arising from proximity are all analogous to transport economies. They arise from the need, for technological or commercial reasons, of rapid communication, often face-to-face with other firms; or from the need to acquire a special service very quickly, as when a machine breaks down.

But the above sort of economies are not, usually, external economies: that is, a firm in an industrial cluster can charge appropriately for the services it renders locally; and it would take account of the benefits it would receive, and pay for, as a result of joining such a cluster, when deciding on its own location (as already seen, the benefit of being able to draw on a trained labour force may be an exception, and constitute the receipt of a genuine external economy). Thus locational problems arise more if one has to design a spatial complex ab initio, for then one has to anticipate and allow for a whole pattern of relationships which need not be ‘external.’ Similar difficulties arise, of course, in the planning of towns, agricultural settlements, and even buildings.

External Considerations Affecting Infrastructural Projects

As already mentioned, the ordinary outputs of many infrastructural enterprises are supplied either free or at prices which are clearly not intended to reflect either cost or benefit. If one calls everything an externality where a difference arises between benefits and receipts, then, for instance, it follows that one classifies all the benefits of a free road system as external effects. The terminology does not matter very much: what matters is simply that every important effect is envisaged and, if possible, valued. The ordinary purpose of a transport project is transport: and the evaluator is not going to forget, in evaluating a road, such things as reduction in traffic costs, time savings, and accident prevention (however difficult it may be to put a value on them, in the absence of a market); in estimating demand he also has to consider the generation of new traffic, and consider the effects on the existing network. It is more likely that things which have nothing to do with transport will be neglected—e.g., noise, or harm or improvement to the landscape. Admittedly, such neglect may seem implausible now that there is so much emphasis on the environment, but it has certainly happened in the past.

Great projects, involving large areas of land and changes in water use, may have major ecological effects which are unforeseen by the planners. This, for instance, is said to be true of the most recent Aswan dam. Similarly, some people fear that the great extension of irrigation may in the longer run have severe diseconomies through increasing soil salinity. Envisaging such effects, which may be better called unintended than external, is outside the realm of economics; although, once their probability is established, the economist may be called upon to appraise them.

Public Goods

So far we have considered only the production and consumption of goods and services that have a market price. But what if a commodity or service—such as national defense or the provision of public information or law enforcement or an antimalaria campaign—is provided to all consumers whenever it is provided to one? If there is this jointness in demand, “free riders” cannot be excluded from enjoying the benefits of the productive activity. A price cannot be claimed from each user. These are cases of public goods or collective goods. A public good cannot be subdivided into purchasable units that can be competitively sold separately to different individuals. Moreover, because a public good (such as clean
air) is supplied collectively to a group and is independent of the number of users, a greater consumption (of clean air) by one individual will not reduce the amount (of clean air) available to another user.

If, however, there is no market for a public good, how is its price to be determined? What should be the public price for a public good? This raises issues of optimal pricing, to be discussed in the next chapter.

At this point we need only note that the collective consumption of public goods is a common instance of market failure, especially significant for a developing country where the provision of public services tends to constitute a large part of domestic output. Essential to health and productivity are sanitation and safe water supplies—public goods that cannot be privately purchased. Agricultural research, the provision of public information, and education are also public goods that are essential to the development process.

The mention of education, however, indicates that there may be different degrees of "publicness" and "privateness" to a good or service. The pure public good merges into other goods and services provided by public expenditures. This is explained in Selection 3.10.

3.10 The Nature of Public Goods

What leads to the choice of public provision from among alternative means of meeting particular ends? Casual observation suggests that the public interest may be served by providing or encouraging provision of a variety of goods or services, and by nonprovision or discouragement of others. The former goods and services may be loosely defined as vested with the public interest, or as public goods.

The Nature of Public Goods

Serving the public interest may take many forms: provision of goods, subsidization of their provision by private means, enactment of laws that impede or prohibit their provision or constrain the form in which they are provided. The focus here will be on policies that involve public expenditures. It should be remembered that choice among alternative available means is always an element in important policy issues.

The goods and services provided by public expenditures or encouraged by public policies can be described and classified in a number of different ways. Though "public goods" are spoken of as a single category, any review of public policies shows great diversity. Some kinds of public goods are provided only publicly because there exists no reasonable private alternative way of providing them. This can happen (as in the case of national parks, national defense, or space research) because there is no private mechanism to pay for these goods, or it can happen (as in the case of sewage disposal or justice) because compulsory use of the good by all is required to permit its enjoyment by any group. Other public goods, such as public housing or public education, may be functionally similar to available private alternatives, but qualitatively different in ways that society prefers. Still other public goods may differ from private ones only in the distribution of beneficiaries and costs.

If the proper domain of public expenditure policy is public goods, their definition becomes vital. The concept has been defined in many ways, and for diverse purposes, and it is not surprising that definitions motivated by purposes other than understanding the rationale and process of public expenditure policy are not wholly satisfactory.

"A public good is any good or service which is de facto provided for or subsidized through government budget finance."15 This definition is neat but not helpful. It does not name the attributes of a potential good or service which a policymaker should take into account in deciding whether to provide the goods. In this definition "publicness" is wholly a matter of legislative designation, not of any characteristic of the good or service.

In many ways, an intrinsic definition based upon technical characteristics of goods or services would be desirable. The perfect collective consumption good


provides such a definition. An impressive array of economists have so defined public goods. Dorfman has recently written:

There are certain goods that have the peculiarity that once they are available no one can be precluded from enjoying them whether he contributed to their provision or not. These are the public goods. Law and order is an example, and there are many others too familiar to make further exemplification worth while. Their essential characteristic is that they are enjoyed but not consumed, [and that their benefits are derived] without any act of appropriation. 17

This kind of very narrow definition was designed to demonstrate that there may be a type of activity that is socially desirable but that will not be achieved by the private market unaided. It serves well the purpose of showing the existence of public goods. It can prove a hindrance, however, if it leads to the view that goods which conform to it are the only class of goods which government can legitimately provide. In fact, examples are hard to find, and the great bulk of nondefense public expenditures covers goods and services that do not meet the definition. Roads, schools, welfare payments, recreational facilities, housing, public power, and irrigation, among others, are important classes of public expenditures that some can be precluded from enjoying, that can be consumed in whole or in part, and that technically can be made subject to user charges. The concept of the perfect collective consumption good, while sufficient to justify public expenditure, is not necessary nor does it embrace much of what public expenditure policy concerns. It does, however, identify certain characteristics such as nonconsumption, nonappropriation, and the existence of externalities that may give a good its public goods aspect.

Externalities are very important, as has been recognized for a long time. While they are important, it would be easy to follow externalities too far. Does any good which produces an externality become a public good no matter how incidental the externality? Few goods do not meet this test and thus this definition fails to provide guidance as to which goods ought to be candidates for public provision, just as it fails to explain which goods are publicly provided.

Since standard definitions are not fully satisfactory, let me try a new one: Any publicly induced or provided collective good is a public good. A collective good in this definition is not necessarily a collective consumption good. Collective goods arise whenever some segment of the public collectively wants and is prepared to pay for a bundle of goods and services other than what the unhampered market will produce. A collective good thus requires that there be (1) an appreciable difference in either quantity or quality between it and the alternative the private market would produce, and (2) a viable demand for the difference.

Collective goods may be privately or publicly provided. Cooperatives, unions, vigilante organizations, country clubs, car pools, and trade associations are all examples of private organizations that arise in response to collective demands for private collective goods or services. When the coordinating mechanisms for providing a collective good invoke the powers of the state, it is here defined as a public good. This definition requires that a public good meet the tests of a collective good. Notice that public provision by itself does not create public goods.

A most important aspect of this definition is that it makes publicness not an all-or-nothing attribute, but one that may apply merely to particular aspects of a good. While there are cases (for example, national defense) in which the choice is between public provision and no provision, and it is thus argued that the good is entirely a public good, the more common situation is for goods to comprise a variety of services, only some of which have the attributes of collective goods. Given sufficient importance, these aspects of collective interest may lead to public provision of either these aspects or the entire good, including its noncollective aspects. Thus provision of smog control or river purification attacks a particular externality of private production. In contrast, public housing provides individuals with services they would otherwise have purchased privately, along with the distinctive public services that public housing is supposed to entail.

Such mixed goods test and stretch definitions. Public education and public housing reflect both quantitative and qualitative differences from the comparable privately produced or producible goods. If the differences are intended and desired, they constitute public goods in this definition.

This somewhat vague notion of public goods can
be filled out by a more detailed classification of different types of public goods.

A Classification of Public Goods

I have, in effect, defined the vector of public goods as a vector of differences between the goods and services the private economy is motivated to provide and the goods and services the public wants, is willing to pay for, and expects its government to assist it in achieving. This is, to an important degree, a normative definition, and much of the debate about the appropriate elements of the public goods vector is a normative debate. But there is a positive aspect as well: What is it about particular goods and services that makes them candidates for public consideration? What is it that makes certain activities the traditional province of governments?

It seems worthwhile to distinguish three types of public goods: (1) those arising from intrinsic, perhaps technical, characteristics of specific goods that result in externalities that are not effectively marketed; (2) those arising from imperfections in market mechanisms, rather than in the nature of the goods or services themselves; (3) those arising from concern with the quality or nature of the environment rather than aspects of particular goods or markets. These become, in order, increasingly elusive, but it is impossible to capture the flavor of actual government expenditure programs without all of them.

PUBLIC GOODS ARISING FROM NONMARKETABLE SERVICES OF PARTICULAR GOODS

The precondition for a discrepancy between public wants and private supplies lies in the existence of externalities (or, as they are sometimes called, spillovers or third-party effects). Any time provision of a good or service yields side effects the value of which is not reflected in the prices of the outputs sold or the resources used, external economies or diseconomies are produced. There can be many reasons for such externalities: private producers may use resources they do not consider scarce, or produce by-products that they do not consider valuable because they cannot control and market them. Familiar examples are discharges of noxious wastes into water or air; downstream navigational or flood control consequences of a private power dam; civic beautification or uglification incident upon building of private golf courses, factories, or slaughterhouses. Because some of the resources used or outputs produced are not correctly valued by the market, there is every reason to expect the market to misuse them. Thus, simply for efficiency's sake, collective concern and public action may be required to allocate resources in accord with true valuations. Whether such externalities (which must surely be present to some extent in every good) justify public notice and action depends upon the benefits to be achieved measured against the costs of interference. People will disagree about the costs of interference or the proper cutoff level; but these are matters different from the nature and size of the externalities.

The perfect collective consumption good is really an extreme case of externalities: All of the output is regarded as individually unmarketable; all of the benefits are external. The outputs of those goods from which one cannot be excluded as a consumer—and thus for which one cannot be compelled to pay his share of the cost of provision—play a large role in the thinking of those who have been concerned to derive a legitimate role for public activity. Defense, public health, law and order, and hurricane watches are familiar examples. The common practice of listing a few examples (and not pressing them very hard) and adding, "There are many other examples" is close to fraudulent. If nonexcludability implies no one can conceivably be excluded, the list of such goods is short indeed. One need not police the ghetto nor defend Alaska. Television signals can be scrambled so as to exclude those who will not buy the unscrambler. Movie houses, concert halls, hospitals, and colleges all use walls to exclude those who will not meet the requirements placed upon their use.

Collective goods may arise because exclusion would be relatively costly rather than because it is impossible. If at any moment this cost is above a certain level, there may be no effective private supply of the good. But in other cases the cost of exclusion may be annoying rather than prohibitive and potential consumers may urge public action merely to avoid bearing the costs. But differently, the cost of arranging exclusion may be an unavoidable externality.

Implicit in this discussion is an important attribute of the public collective good: the willingness to appeal to the police power of the state. One can slide in imperceptible steps from situations where there is no viable alternative means of providing the good, to cases where the alternative seems unnecessarily costly, to cases where the alternative, while not very costly, is simply judged to be less desirable, and finally to cases where the alternative differs only in who pays for it.

There is real purpose in downgrading the distinction between inability and unwillingness to provide a good privately. If a practical definition of specific collective consumption goods and services is to be established, it seems difficult to escape the view that a judgment is required about reasons for turning to the political process and the coercive power of the state, rather than dealing with the second-best solution. These reasons must be judged meritorious by the social decision processes. If this is so, collective
consumption goods are defined by, as much as they define, the exercise of legitimate governmental decision processes.

Among the positive issues that underlie the normative debate about whether a particular collective good ought to be publicly provided are (1) whether private market alternatives to public provision are impossible, impractical, merely costly, or simply unwanted; (2) why the market solution is unsatisfactory to members of the group and to society as a whole; and (3) what the identity of the group of beneficiaries is. The last deserves comment.

A collective good need not provide joint benefits to all members of a society, only to some subgroup. But which group? The larger the group the more persuasive its demand for public action is likely to be, or (put differently) the less willing its members will be to accept a costly alternative. There are bases other than size for weighing the merits of the demands of any group, and these may vary over time. Importers, farmers, labor unions, small businessmen, and minority groups are among the identifiable groups that have asked and received special treatment. Today, for example, our society seems more responsive to the demands of the underprivileged than to those of the wealthy; a half century ago it was clearly otherwise.

One reason many collective consumption goods lead to demands for public provision is because the potential willingness to pay of different consumers cannot be tapped by private suppliers. Weisbrod [has suggested] a further source of values for which there is no market: option demands. Consider several examples. I value the existence of Yellowstone Park, despite the hope that I never have to visit it again; I value the Everglades because I may want to visit them, even though I probably will not. Similarly, I value a first-rate tuberculosis sanatorium, although in all probability I shall never need its services. Were any of these threatened with extinction I should be the loser, but there is no market in which my willingness to pay for the option to use them can be translated into revenue to the providers.

Weisbrod's most suggestive example concerns the standby availability of transport. How much is it worth to the New York-Washington air travelers to have a good rail alternative in case of snow or strike? Suppose that it is worth enough to justify the rail service, but that the railroad has no way of being reimbursed by those whose option demands are critical to continuation of the services. In these circumstances, the public good may be provided by the government's insistence that the railroad's passenger service be maintained with or without subsidy. In this view governments may not have been irrational in trying to preserve passenger train service even in the face of the inability of the carriers to develop a set of user charges that succeed in covering costs.

PUBLIC GOODS ARISING FROM MARKET IMPERFECTIONS

In practice there can be no sharp distinction between market failure caused by technical characteristics of particular goods and market failure caused by market imperfections. Inability to handle externalities, for example, may be regarded as a shortcoming of existing markets rather than as the absence of markets for specific services. But a distinction suggests additional sources of unsatisfactory private market performance that generate demand for public collective action. Efficient markets frequently presuppose adequate information, timely adjustment, sufficient competition, and modest transaction costs. The absence of any of these may motivate replacement of market determination by nonmarket provision, or supplementation of markets with ancillary public goods.

Information. Suppose all conditions for ideal resource allocation are satisfied except that market signals are systematically not read or are misperceived by economic actors. An allocation of goods and resources will occur, but it will, in general, differ from an allocation based on adequate and accurate information. Information may be a collective good (and thus generate a demand for its public provision) because even if there is a well-articulated private desire to have information, there may be no effective market in which to buy it efficiently. It may also be a public good because the externalities of having misinformed traders may be judged to be socially undesirable.

Time lags. If resources respond to market signals surely but slowly, the market process may prove an expensive way to achieve resource shifts. If physicists are in short supply, their price may be expected to rise and this may motivate additional youngsters to undertake education leading to careers as physicists. Since education is a slow process, available physicists may earn high rents over long periods due to the long supply lags. It may well be that public policy can increase the supply of physicists more quickly and more cheaply by fellowships, research grants, and other means than the unaided market. If increases in the supply of physicists, but not increased incomes of existing physicists, are desired results, then such programs supply public goods.

A large and growing literature is concerned with the extent and causes of factor immobility. Education is but one of the sources; others include unemployment rates, prejudice, and institutional barriers to greater mobility, such as seniority and pension rules and state laws affecting eligibility for relief. Whenever markets work to reallocate resources too slowly,
a collective demand to supplement or to replace the market mechanism may arise. Retraining programs, moving allowances, public employment services, and even attacks on prejudice may be public goods if they serve to reduce the lags that the market economy produces to the point that society finds tolerable.

Monopoly power. Noncompetitive imperfections require little comment. Public activities to encourage or compel competitive behavior, or to replace monopolistic, private supply by public provision, are further sources of public goods.

Transaction costs. It has been seen that an important aspect of collective goods concerns the inability of the market to translate potential willingness to pay into revenues. Related is the situation where the private market is technically able to collect revenues, but at a high cost. Toll collection on interurban roads and urban bridges may or may not be both feasible and efficient, but intra-urban toll roads would surely involve intolerably large collection costs and time losses. Because the transaction costs of high speed intra-urban travel as a private good are prohibitive, if it is to exist at all it must be a public good. Metering costs may be justified for commodities of high unit value, such as gas and electricity, but not for sewage (and, in some areas of high population density, for water).

Where these high transaction costs inhere in the particular service they are simply an externality; where they reflect the institutional arrangements of the market they are a potential additional source of collective concern. The higher cost of attempting to gear a pricing system to an individual's willingness to pay is a repeated source of turning away from the market. Suppose for many goods willingness to pay increases at least proportionally with income. With a few exceptions most private services are not provided on a basis that reflects income, because of the enormous administrative costs that such pricing would entail. If such a basis of payment is appropriate, reliance on the income tax, and thus on state provision, may suddenly appear desirable.

PUBLIC GOODS ARISING OUT OF CONCERN WITH ENVIRONMENTAL QUALITY

Up to this point public goods have been discussed in terms of market failure—failure because of either the absence or the imperfections of private markets. This is the grand tradition of classical economics. But even perfectly functioning markets for all goods and services would not eliminate the desire for market interference. Men may choose to reject market solutions to allocative problems with respect to the distribution of income, the nature or quality of goods produced, or the patterns of consumption that markets produce.

The most compelling examples of collective public goods appear to be national defense, law and order, and public health. What is their particular appeal? Is it that they are collective consumption goods? So is television. The appeal is not in the specific planes, rockets, soldiers, policemen, vaccines, or nurses that are their elements, for each of these can be readily provided as private goods to private users, but rather in the fact that they are part of and condition the environment of the society. Even the criminal who detests the legal framework is affected by it. Looked at this way these goods suggest other things that affect the environment and thus create externalities not linked to particular goods: the literacy rate, the level of unemployment, the incidence of crime, the pace of technological progress, and, importantly, the pattern of distribution of income and wealth.

Distribution of income. Accept this assertion: it is fully feasible to charge users for use of parks and playgrounds, to charge parents for school bus service and school lunches, to charge fishermen for fishing privileges. Suppose in each of these cases that there is sufficient willingness to pay and ability to collect to ensure private provision of parks, playgrounds, school buses and lunches, and fishing opportunities. Should these functions be left to private provision?

There are two issues here rather than one. Does concern focus merely on the distribution of income or on the pattern of consumption? When subsidized public housing is provided to the urban poor, is the aim to make available more or better housing to users who would be excluded by private provision (or who would exclude themselves)? Or is it rather to increase their share of national consumption, and the choice of giving them public housing instead of a cash income supplement is motivated by some other consideration? (One might use indirect means, for example, in order not to impair the self-respect of the recipients.)

It is sometimes argued that purely redistributional objectives which reflect dissatisfaction with the initial patterns of ownership of wealth and resources ought to be satisfied by income transfers rather than by provision of goods and services, in order not to distort resource allocation. This familiar argument does not persuade, if one regards as legitimate a desire of a society to interfere with the pattern of consumption that would result from market determinations. A society may choose to affect income distribution and the pattern of consumption jointly. Provision of housing, education, milk, or recreation to underprivileged children may be a public good because of the externalities which children so treated
bestow upon others. Public policies designed to aid small business, the family farm, the needy aged, and the slum child all reflect rejection of market determination, rather than denial of the possibility of market determination.

It is, of course, not clear that all actual interferences reflect a positive intention both to redistribute income and to change consumption patterns. In the United Kingdom (by way of contrast with the United States), fishing rights are sold, and fishing is an upper-class form of recreation. On the other hand, virtually all Scottish golf courses are owned by municipalities and subsidized out of tax receipts, and in Scotland golf is a working-class recreation. But if some consumption distortions are fortuitous, others are intended.

Nature and quality of output. The quality and nature of some goods and services are of public concern, quite independent of any distributional considerations. Often the nature of the good or service is affected—for better or for worse—by who provides it. Government newspapers differ from private ones, public television and radio from commercial broadcasting, a system of public schools from a private school system, private from public research and development. In some of these examples the two kinds of goods may coexist; in others an exclusive choice is made. But in all cases a choice among qualitatively different outputs may and can be made: the qualitative difference of public from private provision constitutes a public good or a public bad.

PUBLIC GOODS: A SUMMARY VIEW

I have stressed the pluralistic nature of the sources of collective demands as arising from technical characteristics of particular goods, from market imperfections and failures, and from other divergences between collective and individual values. The time is long since past when the primary need is to define public goods merely in order to establish the prima facie case for some public interference with private markets; what is sought instead is a framework for debate about whether particular activities merit inclusion in the public sector.

It seems to me useful to identify in each case the source of the alleged collective concern. Is the source a major qualitative difference between public and private provision, or is it merely a wish for incremental output, arising in response to a neglected externality? In this distinction often lies an important policy choice between public provision and a less fundamental public restructuring of private incentives. Similarly one wants an indication of whether public concern is fixed on the specific good or service or on the environment in general. There are more ways to reduce overall unemployment than there are ways to retrain Appalachian miners. Again the relevant alternatives are affected by the real objects of policy. Frequently at issue is whether redistributional policies achieved by provision of specific goods and services bring about changes in consumption patterns deliberately or incidentally.

The basis of collective concern having been established, it is worth establishing the basis of public concern. Who are the alleged beneficiaries, and what is their claim to recognition? What second-best alternative do they face if their claims are rejected? Defining as specifically as possible the vector of differences between a private good and its public alternative is a critical part of the public decision-making process. Neither de facto definitions, nor neat but narrow ones, such as that of the perfect collective consumption good, prove very helpful for the crucial problem of defining the scope of the public sector.

The foregoing helps define the scope of the public sector. Even perfect competition therefore fails to achieve the ideal resource allocation unless there are no externalities and the resultant income distribution is considered desirable. For every different income distribution, there will be a different Pareto optimum. For a developing country where the inequality in the distribution of income and absolute poverty are so important, the concepts of perfect competition and Pareto optimality are very limited indeed. The additional questions of equity, or fairness, or distributive justice must also be raised.

Market failure—as evidenced by market imperfections, externalities, public goods in the narrow sense, and income distribution—has enlarged the scope of the public sector and the extension of public services. This effect raises, in turn, fundamental public policy issues. We have considered some of these issues when discussing remedial policies to correct market failure. Beyond these policies, governments frequently turn to the creation of a public enterprise as an alternative to market failure. This then raises the question of how public prices should be established for public products. To this question we turn in the next chapter.
CHAPTER 4
Public Pricing Policy

Given the prominence of public enterprises in their economies, the governments of developing countries must solve the difficult problem of deciding what should be the method of payment for public services. The conventional advice of the economist is that price should equal marginal cost. Such a pricing rule should enable the consumer to indicate the value of an additional unit of output, and thereby provide a guide to investment and project selection. But the rule of marginal cost pricing becomes more complicated—with several possible qualifications—as governments attempt to implement it.

First, measurement problems arise in determining how to calculate short-run marginal cost, long-run marginal cost, and variations in marginal costs at different locations. A strict application of marginal cost pricing may be impossible because of these problems of measurement, but improvement over the existing pricing system may still be possible to achieve a better—even if not optimal—allocation of resources. Moreover, even if temporal and locational variations in costs are determined, the question still remains whether an optimal pricing policy should be restricted to only the objective of efficiency in resource allocation, or should also generate enough revenue to cover costs, and should meet an objective of “fairness” with possibly no user charge or some discriminatory pricing. Under certain conditions, there may also be a case for the granting of subsidies. And pricing policy will inevitably affect investment policy.

Beyond the economic objective of achieving the optimum use of resources (efficiency), the other objectives of public pricing must therefore also be emphasized. It is important to give due weight to the financial aspect of public pricing to ensure not only solvency of the public enterprise but also its ability to generate cash internally and raise in the form of debt and equity capital the resources needed for expansion. With the emphasis on reduction of absolute poverty and fulfillment of basic needs, more attention must also be given to the social aspects of public pricing—that is, to the ways by which public pricing might affect the distribution of income or other social priorities.

Along with the marginal cost pricing rule, this chapter will examine the major qualifications or extensions to this pricing rule. It will be seen repeatedly that a public pricing policy must have at least three dimensions—efficiency, financial cost coverage, and distributional considerations.

The Marginal Cost Pricing Rule

From previous chapters we can understand how marginal cost pricing can be considered an efficiency pricing policy. If the marginal value of an increment of output is indicated by its price, and its marginal cost indicates the forgone op-
portunity of using resources elsewhere, then an output that equates price to marginal cost can be considered "ideal" in the sense of maximizing net economic benefits. No other output would allow an increase in net consumer benefit, and hence the resource allocation is efficient. In applying the marginal cost pricing rule, however, the concept of "cost" is crucial. The true opportunity cost or social cost must be measured by marginal cost.

Although we have repeatedly noted that the equating of marginal cost to price is a condition of efficiency, we must also recognize several qualifications to this principle of public pricing. One problem arises whenever economies of scale are not fully exploited: average costs are then decreasing and marginal cost is lower than average total cost, so that selling output at a price equal to marginal cost yields a total revenue that is below total cost of production. All industries that operate under decreasing cost conditions will encounter such a deficit. Public pricing then confronts the dilemma of how to sell at marginal cost for the sake of efficiency and the best utilization of capacity and yet cover total costs of production. There are various ways of resolving this dilemma, according to different pricing practices and different ways of financing the deficit, as discussed in some of the readings below.

Issues of "fairness" also qualify the principle of marginal cost pricing. In a developing country, it may be questioned at the outset whether consumer willingness to pay for a product or service is the appropriate basis for deciding on the price. The fulfillment of basic needs or a transfer of income to the poorest of consumers may also invalidate a strict application of marginal cost pricing.

The following readings explore the marginal cost pricing rule and its qualifications in practice.

4.1 Principles of Public Pricing

A basic proposition in economics is that prices should equal costs. If the price charged for a product or service is less than the economic cost of producing it, then people will consume more of it than they would if price equaled cost (provided the price elasticity of demand exceeds zero). Additional resources will therefore be used to produce the underpriced good or service. Since the economic cost of using resources to produce an item reflects their value in producing other goods or services, it is clear that the result of this policy will be to divert resources to produce less valuable (less desired) rather than more valued (or desired) goods or services. Conversely, if price is set above cost, too little of this product relative to other products will be produced. Since neither result makes much sense, the correct price must be that which just equals cost. This is the economic rule for the efficient allocation of resources.

More precisely, cost in this argument means marginal cost, defined as that cost which is directly attributable to a small increase or decrease in production. In deciding whether to produce an extra unit of some good or service our concern is with what we give up in terms of not producing some other good. The relevant cost is therefore measured by the additional or marginal cost of producing the extra unit, not the average total cost of production per unit (which also includes sunk or overhead costs which are incurred whether or not the extra unit is produced). For economic efficiency in the allocation of resources, then, price should equal marginal cost. This proposition is as true in principle for public as for private products—and as difficult to apply in practice.

The Case for Pricing

The economic case for pricing whenever possible in the public sector rests essentially on the contribution that pricing, particularly marginal-cost pricing, can make to allocative efficiency. Without prices, public sector managers and planners are deprived of an important objective indicator as to how resources should be used. Correct prices have the great theoretical and practical virtue of providing signals which will, under certain conditions, accurately indicate the

Selection 4.1 reprinted, with permission, from Richard M. Bird, Charging for Public Services: A New Look at an Old Idea (Toronto: Canadian Tax Foundation, December 1976), pp. 33–42.
quantity and quality of things which people desire. Despite the inevitable problems of determining and levying the right prices, a strong case can be made that more use of pricing in the public sector would lessen the number of occasions on which the wrong product is produced, in the wrong quantity, and with inappropriate quality differentiation.

At present there is no mechanism by which demand, or the lack of it, for even those public sector activities which consist of providing essentially private goods can be recorded or can directly influence the allocation of resources in the public sector. When consumers pay for public services through the price system their individual actions can signal shifts in demand more quickly and flexibly than can the inherently cumbersome political mechanism. Through prices, consumers can also be offered differentiations in the quality of public services—whether trash collection, off-street parking, or foreign language instruction—which otherwise can only be offered on an all-or-nothing basis. Provided that prices are set efficiently, and that the limitations on the use of pricing discussed below are not overriding, pricing thus has the inestimable virtue of providing relatively clearcut guidance to decisions on both the quantity and quality of the production of a wide range of public sector activities.

All students of government, as well as practising politicians, are familiar with the constant pleas from every section of the community for the extension of those public services which accrue to their direct benefit—pleas frequently accompanied by criticism of the expansion of government activity in general (that is, for other people's benefit). One economic explanation of this phenomenon is that much of what is done by the government does not provide "public goods" in the economist's sense at all. Instead, much government activity consists of providing goods which are wholly or mainly private in character and then distributing them in some way or another, usually for free. If the supply of a private good is organized by a tax-financed system and is then distributed without direct charge at the point of use, the likely result is a perceived shortage and underinvestment.

In the political arena, shortage is thus largely a function of prices, or their absence. The likely consequence of this situation is constant pressure for more expenditure on the activity, since potential beneficiaries correctly perceive little or no direct cost to themselves in such expansion. The result of this pressure is a misallocation of resources, with too many public-sector resources flowing into the production of freely distributed private goods.

Whenever government is engaged in providing goods which are essentially or largely private in character, charging those who benefit directly would as a rule improve efficiency, promote equity (at least in the benefit and sometimes in the ability sense), and relieve the economic distortions arising from the taxes otherwise needed to finance the activity. These gains may to some extent be offset by the loss of external benefits arising from the consumption of the service (education) or by the inability to achieve the same distributional result in some other way (health). But corrective measures can be taken to overcome these problems, and the main point—that pricing where possible is a sounder general principle than the more common notion that it is somehow wrong for government to sell services—remains untouched. Once the mistake of giving something away has been made, however, it is generally very hard to overcome the interests of the beneficiaries. Nor is there much pressure to do so. Instead, just as with tax concessions, other groups usually clamour for equivalent access to the public trough in the form of expansions of one free service or another.

An important advantage of pricing is thus to curb the demand for expanded public sector activities by making their real costs apparent to the prospective beneficiaries in a meaningful fashion. Correct pricing can alleviate both the pressure to expand government and, by providing additional finance, the shortage of revenues which sometimes, especially at the local level, restricts that expansion, even when warranted. Pricing achieves this double miracle simply by being correct, that is, by correctly reflecting opportunity costs and people's evaluations (as measured by their willingness to pay) of the services in question.¹ If they are not willing to pay what the expansion of a service will cost, then it should not be expanded; if they are, it should be—and financing to do so will be available.

There are thus two principal reasons urging wider use of public pricing and devices such as special benefit taxes. One is the rare possibility of increasing public revenue in an economically efficient way. The other—really the other side of the same coin—is the improvement in the information available to guide the rational allocation of other public sector resources.

Limitations on Pricing

That this apparent cure for some painful dilemmas has not swept the world, however, suggests that there are also powerful arguments against more use of pricing in the public sector. These arguments may be summarized as follows:

1. There is obviously in today's world a basic public mistrust and misunderstanding of the role of prices in general.

1. This formulation obviously sets aside the problem of the "justice" of the initial distribution of power and wealth.
2. There are both inherent limitations to the efficiency case for pricing and technical difficulties in implementing correct pricing.

3. Finally, in both theory and practice the distributive effects of increased reliance on prices may give cause for concern. . . .

In principle, economic efficiency will be achieved by setting prices at marginal cost. In practice, there are many problems in implementing this deceptively simple principle. One set of such problems, for example, concerns the correct definition of cost and output, particularly with respect to public sector products. Other technical problems may arise from the unavailability of needed cost information, the high cost of collecting it, or the importance of such non-monetary costs as waiting time. There are also more basic difficulties such as the importance of goals other than economic efficiency and the lack of public acceptance of charges for government activities. It is neither useful nor possible to review here the many controversies about marginal-cost pricing which have long occupied welfare economists. Instead, this section simply reviews several points emerging from this discussion which are relevant throughout this study. These points range from the theoretical to the severely practical.

One of the more theoretical points is that if there are some sectors of the economy in which marginal-cost pricing is not in force, then it is not correct in terms of efficient resource allocation to implement such pricing in other sectors. Rather, the aim should then be to establish roughly the same deviation between price and cost in all sectors. Since in reality there are large sectors of the economy in which prices do not equal marginal costs, the efficiency case for enforcing strict marginal-cost pricing in any particular sector is inevitably weak. Nevertheless, rough rules of thumb approximating the correct solution can undoubtedly be developed which will produce more efficient results than the most probable alternatives of either zero or full-cost (average cost) pricing. Marginal-cost pricing may be an unattainable ideal, but it is a good deal better guide to allocative efficiency than no ideal at all.

Another area of difficulty with marginal-cost pricing concerns externalities, that is, the benefits accruing to persons other than those who directly benefit from the public service. In some instances, the attention paid to this justification in fact serves to mask the primarily private nature of the benefits. In others, however, it is quite true to say that an individual's consumption of some service (e.g., primary education) may benefit everyone else, at least to some extent. In the public sector the justification for much activity is, of course, the general benefits which are supposed to result.

The mere existence of such external economies does not mean, however, that the goods should not be priced. If there is indeed a social benefit to be derived from individual consumption of certain services, individuals might, for example, be given subsidies to be spent on health, education or housing or whatever, and the allocatively correct price charged. In this way the socially correct amount of the service would be produced (and consumed), whereas zero-pricing would almost certainly lead to overproduction. In practice, however, given the way the political system works, it seems unlikely that correct pricing could be expected in these circumstances. Furthermore, investment decisions could certainly not be governed by financial results alone. Nevertheless, again the correct standard from an economic point of view is still as close an approximation to marginal-cost pricing as possible.

A particular externality problem is that arising from the openness of political jurisdictions. Some of the benefits of services provided in one local government area will accrue in part to the benefit of residents of other jurisdictions, who cannot be directly charged for them. This situation gives rise to various problems of vertical and horizontal fiscal coordination: the former may be resolved in part by a system of intergovernmental grants and the latter, less satisfactorily, by intergovernmental bargaining. Again, however, it is too easy, and wrong, to conclude that the inability to charge perfectly for many services means that one should not try to do the best one can.

While marginal cost can be defined to include certain elements of what are sometimes called fixed costs—for example, when additional output requires an addition to productive capacity—as a rule marginal-cost pricing means short-run marginal-cost (SRMC) pricing. The relevant costs are therefore those which vary when one additional unit is produced. In many cases of publicly provided goods and services the revenues generated by following this rule will not be sufficient to cover total cost. SRMC pricing will then result in an operating deficit, which will in turn have to be financed, usually by taxes which cause their own inefficiencies in resource allocation.

One suggestion which has been made is that the marginal cost (in terms of resource inefficiency) attributable to the need to finance this deficit may be handled, at least in part, by raising prices sufficiently above marginal cost in the first place so that the inefficiency (or marginal cost) of raising revenue in this way will be roughly equivalent to that arising from other sources of financing. In other words, the

2. This is, of course, the traditional case of decreasing average costs of production.
margin to be equated in the efficiency exercise is the marginal cost of raising funds, and there is no presumption that deficits resulting from marginal-cost pricing need be financed from general revenues. They may well be best financed by departing from the strict marginal-cost standard.

While many other points could be mentioned, this discussion may be summed up by noting that pricing is most useful, practical, and acceptable when several conditions are satisfied. First, in order to charge any price it is necessary that those who benefit directly from the service can be excluded from enjoying it if they do not pay. Secondly, most benefits should accrue to the primary recipients of the service so that there is little possibility of loss of significant externalities. Thirdly, the demand for the service should be elastic, so that imposition of a price will affect allocation and hence help achieve the efficiency objective. Fourthly, collection costs (including those of enforcing excludability) should also be low. In order for a strict pricing policy to be feasible, both the quantity and the quality dimensions of output units must be capable of being specified, and the prices must be enforceable and collectable at a reasonable cost. Finally, there should be no unacceptable inequities resulting from the imposition of prices.

In principle, if these conditions are met to an acceptable degree, the price for the service should be set to cover only the marginal costs attributable to the direct benefit accruing to the individual who will have to pay (including, however, any social costs arising from individual actions, such as congestion and pollution). It may also, of course, be necessary to adjust the price to allow for such factors as the deviation from marginal-cost pricing elsewhere in the economy and the marginal cost of raising funds.

The price of pricing and of enforcing excludability, even where it is technically feasible to do so, needs to be particularly kept in mind since in practice it constitutes a major deterrent (along with the distributional aspects of services such as education and peak-hour transit) to extending the scope of public pricing. It seems likely, however, that many of the traditional technical problems in these areas will decline in importance with the advance of technology, just as the development of an adequate income-support system will weaken the distributive case against pricing.

There are, however, less tractable problems in enforcing pricing in the public sector owing to the inevitably political nature of the process. There is some political cost—no less real for being nebulous—in reaching agreement on any pricing strategy. There is also a strong possibility that the combination of these political costs on the one hand and the resistance of most bureaucracies to change on the other will mean that publicly established prices will be too sticky or inflexible—that is, that they will not be modified to cope with changing circumstances, as they should be in the interests of efficient resource allocation.

**Alternatives to Marginal-Cost Pricing**

A reasonable general conclusion to this point is that considerable attention should be paid to short-run marginal costs (SRMC) but that in most instances prices will probably best not be set exactly equal to SRMC. As noted earlier, the basic rationale for introducing pricing into the public sector where possible is really the same as that underlying Adam Smith's famous "invisible hand" doctrine, that is, to achieve an efficient allocation of resources by having each individual choose to consume that amount he wants when offered at a price which reflects the value (in terms of what else they could produce) of the resources used in producing it. Unless such a price is charged, consumers will demand more and more of the service to the point at which an additional unit of the service is worth nothing to them, regardless of the social cost of doing so. Unless individuals bear the full cost of their actions, society's resources will not be allocated so as to yield the maximum possible satisfaction to each individual. Only marginal-cost pricing will accurately measure the costs and therefore only marginal-cost pricing will serve this purpose. Since in practice, however, marginal-cost prices cannot always be achieved, or other objectives may require that different prices should be charged, this ideal position cannot always be attained. Nevertheless, it remains a viable and useful goal at which to aim.

In addition to the modified marginal-cost rules suggested above, a number of other compromises to cope with some of the real world difficulties with marginal-cost pricing have been put forward. Perhaps the most commonly suggested alternative is a multi-part tariff. In this scheme, the consumer of the service pays a certain fixed charge for access or connection to the service and then an incremental charge related directly to the amount of the service which he uses. This incremental charge should, of course, be related to the marginal cost of providing the service to the consumer, while the inframarginal residue is to be covered by the fixed charge, which should be designed so as not to affect the level of use.1

---

1. In other words, the demand for access to the service is generally assumed to be less elastic than the demand for use of the service. It is quite possible, however, that the "connection fee" will exert an undue deterrent effect on access, in which case it may be better to have the use price above SRMC with a lower fee in order to induce marginal consumers to stay in the system and pay the license fee.
Although this technique is already used to some extent in practice, particularly by some public utilities, it is seldom fully applied and could be used much more widely.

Although in principle the optimality of this device depends on the unlikely matching of the fixed charge paid by each person with the consumer surplus he reaps from the service, in practice a much rougher degree of matching (e.g., by income levels) might produce results superior to the prices which would otherwise be set. On the whole, variants of the two-part tariff approach look very attractive in economic terms as a means of pricing a wide range of public sector activities. One such variant recently proposed for a number of local government services amounts to a three-part tariff: a use price based on the marginal cost of consumption of the service (e.g., water), a lump-sum access charge (a connection fee, or a local tax), and a distribution charge varying by location (equal to the long-run cost of providing facilities for delivery at site of the demanded volume). The third of these pricing components could be levied by a special assessment or development charge.

A more common approach to pricing public services is traditional “full” or average-cost pricing. In view of the constraints on the implementation of correct marginal-cost pricing noted above, it has been argued that this solution may not turn out to be as allocatively bad as it has sometimes been painted. Others have stated, baldly, that “average cost pricing is unambiguously inferior to all other alternatives by all criteria” (referring to a particular study of pricing certain local services). The safest conclusion is, as usual, that there is no absolutely good or absolutely bad system but only a choice of imperfect possibilities. Which system is better in a particular instance is fundamentally an empirical question, though unfortunately one where we simply do not have the relevant information.

Much the same can be said of another traditional device often maligned by economists, namely, queuing in lieu of price rationing. Rationing by time rather than by money is usually an inefficient procedure. It may, however, in a sense be considered to be an equitable one, since time is more equally distributed than money. In addition, the money cost of time is presumably higher for the rich than for the poor. A pricing system combining time and money prices might therefore produce both efficient and equitable results in certain instances. Those with more time than money could wait in line, while those subject to the opposite constraint could pay to be advanced in the queue. Various time-price pairings in effect would create a degree of product differentiation which would enable the supplying authority to match the prices charged particular individuals with the consumer surpluses they reap from consuming the service in question. This intriguing result assumes that it is politically and administratively feasible to have such differential pricing, which may be far from the truth.

With respect to these alternatives to marginal-cost pricing, generalizations are almost certain to be wrong. What is needed is a careful case-by-case study of the implications of alternative pricing strategies for different public sector activities. Such an evaluation of alternative pricing possibilities should also be made a part of every reputable cost-benefit study of the attractiveness of this or that investment, in lieu of the all too common practice of simply accepting whatever pricing system exists (or does not exist) and basing on it all the elaborate projections and calculations which are made in such analyses. Eclectic pragmatism ought therefore to be the rule in this as in many other areas of public policy, as should attempts to introduce more flexibility and experimentation.

Any practical solution to the pricing problem will necessarily distort the optimal efficiency conditions. Each case must therefore be considered separately, and no general rule can be laid down. It is, for example, dangerous to stress efficiency exclusively since this implies acceptance of the present income distribution. It is fine to say that any distributional problem can be resolved by the tax-transfer system, but it would be well to ascertain just how the system functions before relying on it too blindly.

Despite these various qualifications and cautions, however, it deserves to be reiterated in conclusion that only correct marginal-cost pricing can provide the informational and efficiency benefits stressed above. In a sense, therefore, such pricing ought to remain the goal, no matter how much it has to be compromised in practice. To the objection that economists tend to overdo the importance of the allocation problem when they make statements such as this, the reply has to be that if the economist does not stress efficiency, no one else is likely to do so either. While the other objectives with which those who set public prices are charged—distribution, regional balance, or whatever—are most unlikely to be forgotten, since each has its fervent advocates, efficiency very likely will be forgotten, and in fact often has been.

The economist dealing with public pricing may sometimes seem to be a rather strident advocate of some esoteric doctrine known as marginal-cost pricing who ignores everything that really matters. In fact, however, he is likely to be a lone voice for
efficiency in a chorus of shouts for inefficiency in order to achieve this or that desired social or political goal. The stridency of his tone may reflect his need to shout also in order to be heard. The problem is not that efficiency is, to the economist, all that matters: it is rather that efficiency does matter, along with other things, and that the best way we know to get efficiency is by setting prices equal to marginal costs. Moreover, no society can do everything that everyone wants done, and so long as resources are scarce, increasing the efficiency with which they are used in any particular activity frees some resources for use elsewhere. Efficiency, then, is no mean virtue, and the economic path to improved efficiency through correct pricing deserves careful attention whenever conceivably applicable and, it is to be hoped, application whenever practicable.

4.2 Marginal Cost Pricing in Practice: The Case of a Bridge

The simplicity of the rationale just given for the marginal cost pricing rule is deceptive because it conceals three important sets of difficulties. To explore these difficulties it is useful to work with a specific example. A classic in the economics literature is the problem of the appropriate pricing policy for a bridge and so we adopt this as our example.

Marginal Cost Pricing in Practice: The Case of a Bridge

It is proposed to build a road bridge across an estuary linking two towns. The problem is to determine whether the bridge should be built and, if so, to choose its capacity, measured in vehicle trips per day, and a toll per vehicle.

The marginal cost pricing principle suggests the following approach to the problem. First estimate the relationship between the level of toll and the number of trips per day, for each day from the date at which the bridge would be available. We assume this demand relationship will remain the same from day to day—no growth or seasonal variation takes place. Next estimate the total costs of constructing and operating a range of bridges of different daily capacities. These costs can be easily put on a daily basis: the interest on the capital expenditure involved in constructing the bridge can be expressed as a cost per day, as can maintenance costs, lighting, toll collection, etc. The larger the capacity of the bridge the greater will be the total daily cost. We make the simplifying assumption that this total cost increases proportionately with capacity, so that a doubling of the latter implies doubling the former. Moreover, at the planning stage, the capacity of the bridge can be varied as finely as we like, so that a bridge of, say, capacity 999 trips per day can be compared with one of capacity 1,000 trips per day.

Now suppose we find that:

(a) at a toll of 50p per vehicle, it is estimated that 1,000 vehicles per day will use the bridge;
(b) the increase in total costs per day—interest, maintenance, etc.—in going from a 999-trip-capacity bridge to a 1,000-trip-capacity bridge is 50p, so that this is the "long-run marginal cost" of a trip.

Then the marginal cost pricing solution is to construct a bridge of 1,000-trip capacity and plan to charge a price of 50p per vehicle. The reasoning is as before: this choice of price and capacity will ensure that the value of the marginal trip across the bridge is just equal to the increment in cost required to supply it. Note also that a result of the assumption that total costs increase proportionately with capacity is the equality of marginal cost with average cost. So at a capacity of 1,000 trips per day total cost is £500 as is total revenue—the bridge will break even. Finally, since the total value of the bridge to consumers exceeds their total expenditure of £500 per day (because in general the price a consumer would pay rather than go without a good entirely exceeds the price he does pay, i.e., there is a "consumers' surplus") while total costs are £500 per day, total benefits exceed total costs and construction of the bridge at the optimal capacity is justified.

Note that in general there are two steps in the economic decision: to find the optimal price and capacity, which is essentially a matter of weighing up costs and benefits at the margin; and to find whether the investment is worth carrying out at all, which involves a comparison of total costs with total benefits. One of these steps can be omitted only if some special assumption is made about its outcome, e.g., the first step will be redundant if only one scale of operation is feasible.

We can now confront this text-book solution with examples of the three sets of problems referred to earlier.

Suppose, however, as is quite possible, demand had been greatly overestimated. At a toll of 50p the bridge is working at, say, only half capacity, while even at a zero toll fewer than 1,000 trips would be made. If the true demand had been known a smaller bridge could have been built, but that is now by the way. The bridge exists, its capital costs are sunk costs, and the only costs which now vary with the number of trips are maintenance costs.

There now appear to be two marginal costs, the "long-run marginal cost" of 50p which incorporates capacity charges, and the "short-run marginal cost," consisting of the extra maintenance costs arising from the increased wear-and-tear of an extra bridge-user, and containing no capital charges because these are fixed costs.

This appearance is however an illusion. There is only one marginal cost, that corresponding to the actual change in total costs as the number of trips varies, and this of course is the short-run marginal cost since the interest costs are fixed and do not vary with utilisation of the bridge. The long-run marginal cost was only relevant at the planning stage, when capital costs could be treated as variable. Given the level of demand which has actually resulted, application of the marginal cost pricing principle implies a price below 50p. If 50p per vehicle were charged, while, because demand is well below capacity an additional trip would impose an extra cost of, say, 5p in increased maintenance costs, then net consumer benefit is increased by reducing price to 5p and increasing utilisation of the bridge. Thus we have "short-run marginal cost pricing."

The general conclusion is that the question "short-run or long-run marginal cost pricing" is essentially bogus. We are in actuality always in the short run when we have to set current output and prices, since in the current period capacity is always fixed. Long-run marginal costs are relevant only to planning and the planned prices will turn out to be appropriate (on the grounds of marginal cost pricing) only if demand is correctly forecast.

The example, however, raises a number of important issues. By setting the toll at 5p the bridge will incur heavy losses. In effect, its capital debt is being written off because of forecasting errors, and this is open to two objections.

First, it creates a problem of management control. It is often argued that "appraisal optimism" is a characteristic feature of public sector investment decision-taking. This is a polite term for the situation in which those people concerned with appraising a project for some reason want it or think it should be carried out, and are prepared to fit whatever numbers are required into the cost and benefit columns to ensure that it meets the criteria currently applied. The approach to pricing just set out appears to con-
make a net contribution to the Exchequer as part of the general system of indirect taxation, though for institutional reasons this contribution is usually expressed as a "financial target."

We can interpret this argument as introducing an objective into the determination of nationalised industry prices additional to that of allocative efficiency. We could call this objective "profitability" as long as this is understood to mean "net contribution to the Exchequer" and not "net income available for distribution to shareholders." This then leads to the next major problem which faces the "marginal cost pricing doctrine."

(II) OTHER OBJECTIVES

In White Papers on the nationalised industries it is usual to see emphasized the point that economic efficiency is not the only goal: the pursuit of this will be subject to considerations of the "wider national interest." This latter form of words is, probably by design, vague and under-specified, and helps neither the industries' boards nor academic observers to determine what the objectives of the industries actually are. It is useful therefore to reduce the area of uncertainty by trying to specify more precisely the dimensions of the "wider national interest."

By examining the history of ministerial intervention in the decisions of the industries it is possible to say that there are, in addition to economic efficiency, three further objectives which government, usually implicitly, wants the industries to promote.

Profitability

The question of the net Exchequer contribution of the bridge arose quite naturally in discussing the problem of pricing when demand had been overestimated. The existence of profitability as an objective, however, implies that the financial out-turn for the bridge must be considered at the planning stage, before it is built. On the earlier assumptions the bridge of capacity 1,000 trips per day would break even at a marginal cost price of 50p, with expected daily revenue and costs of £500. Suppose, however, that in line with the general policy of indirect taxation, the bridge is set a profit target of £50 a day (which may then be expressed as a rate of return on assets).

Now if demand were independent of price the profit target could be met by building a bridge of capacity 1,000 trips a day and charging a toll of 55p. More realistically an increase in price would reduce demand; so assume that a 10 percent price rise would imply 5 percent, or 50, fewer trips per day. A 1,000-trips-capacity bridge would be operating below capacity and falling short of its profit target by £25.00 per day (assuming that each trip not made saves 5p in maintenance costs). The existence of a profit target implies a smaller bridge as well as a higher price. The target will roughly be met if a bridge of 950-trips-capacity is built and a toll of 55p charged. The general point is of course that the imposition of a profit target will lead to a departure from the marginal cost pricing solution except in the special case in which the "target" is actually derived from that solution. In this latter case the bridge would be set a profit target of zero; it would be required simply to break even.

Income Distribution

Many instances of intervention by government to make or amend nationalised industry decisions are designed to further the interests of well-defined groups in society—groups of consumers in a specific area and/or income group or of workers in a particular occupation. The essential characteristic of these interventions in economic terms is that they make the real incomes of the groups higher than they otherwise would be, and therefore the real incomes of non-members of those groups rather lower—they are essentially ways of re-distributing real incomes. We can use our bridge example to show how the income-distributional preferences of government may affect nationalised industry pricing policies.

The planners have arrived at the marginal cost pricing solution (the profit target is just to break even) of a 1,000-trip-capacity bridge and a toll of 50p, but suppose that the relevant minister finds this toll unacceptably high—it will put use of the bridge out of reach of poorer members of the population and specifies a maximum toll of 25p. Assuming now that this 50 percent reduction increases demand also by 50 percent, this will imply that the bridge must be built to a capacity of 1,500 trips per day (to set the lower toll and keep capacity at 1,000 would involve enormous congestion and delays to users). Since the total daily cost of the bridge will be £750 (50p is the long-run average = marginal cost per trip) while daily revenue will be £375, the bridge authority must budget for a daily loss of £375.

The beneficiaries of the minister's intervention are first of all those who would have paid 50p for the trip—their real incomes are now higher by 25p for each trip they make—and also those who will now make the trip when they would not previously have done so. The value of the benefits to the latter group,

5. Actually there is a shortfall of £2.50 per day at this toll, but going up by a whole penny to 56p overfulfills by almost £2 per day. The government must choose!
i.e., their consumer surplus, can be estimated roughly at £62 per day. Thus we could put the real income gains to the bridge users at £312 per day.

Someone, however, has to pay—the loss of £375 has to be financed. If the bridge authority supplies other goods and services—perhaps operates other bridges—then it could plan to meet its break-even target by raising their prices, thus reducing their consumers' real incomes. Alternatively, the bridge may be subsidised (e.g., by writing-off debt) and therefore the loss will effectively be financed by increasing taxation or reducing public expenditure generally. Thus the income redistribution is from taxpayers or beneficiaries of public expenditure in general to the users of the bridge.

The income-distributional goal is clearly being pursued at the expense of the "profitability" goal in this example. It also represents a sacrifice of economic efficiency. To see this, note that the cost of the extra resources required to generate the 500 additional trips is £250 per day—in other words output worth that amount has to be sacrificed elsewhere in the economy. The value of those extra trips to the people making them is the amount they pay, £125, plus the consumer surplus of £62. Thus there is a "deadweight loss" to the economy, the excess of value of resources over value of benefits, of £63—the economy has irrevocably lost consumption benefits of that amount as a result of the policy of income redistribution.

This deadweight loss is part of the reason economists often argue against the use of nationalised industry prices to redistribute real incomes. Another is that it may be an inefficient way of achieving the redistribution per se. If the bulk of the bridge users are not among the group whose real incomes it is desired to raise, then a large part of the real income gain of £312 per day is going to the wrong people (note that in the present example only about 20 percent of the gain accrues to the people who would not have made the trip at the higher price). Moreover at least some of the taxes or expenditure reductions required to finance the loss may also be coming from the wrong people. The minister's desire to see that the less well-off can use the bridge would be better served either by raising their incomes through the social security system, or by devising some means of giving discounts on the toll only to them, e.g., by the presentation of a special card at the toll booth. Of these the former is probably to be preferred—it has the added advantage that the benefits of the increased real income do not have to be taken in the form of trips across the bridge.

4.3 Marginal Cost Pricing

Why sell at cost? More precisely, why sell at marginal cost? What marginal cost is involved? These are the three themes which will be developed in this short discussion.

Why Sell at Cost?

Let us make clear at the outset that we are not referring to the existence of a profit margin when we talk about sale at cost. Provided that it remains reasonable—and no one would think of maintaining that profit should be unreasonable—the margin of profit represents reward for the capital employed in the business, and compensation for its management. By this definition, the profit margin is an element in cost, in the sense in which we are using the term here, as much as wages or fixed financial charges (payment to bondholders and other creditors). The nature of a "normal" return to capital and management could doubtless be discussed; but this discussion might just as well be about normal levels of wages or the rate of interest. In any case, this discussion is not our object here.

Assuming that cost has been defined in the broad sense being considered here, the problem posed is: Can sale above cost to some consumers to finance sale below cost to others be justified, or must sales to all be at cost?

"To govern is to choose." This adage is as appropriate for the head of a firm as for governmental Cabinets. The life of the head of a firm is made up of daily choices; those choices which bear on the economic aspects of his activity are directed by costs. For an equal service rendered, the head of the enterprise chooses the least-cost solution.

If the prices of the supplies he is thinking of using, and in particular of his supplies of electricity, are equal to their costs of production, the least-cost solution for him is also the least-cost solution for society. Otherwise the least-cost solution for the enterprise may be an ex-

6. A standard way of measuring the benefits of the "generated traffic" is by the area of the triangle whose vertical side is the price reduction and whose horizontal side is the demand increase. In the present case this measure is £1\frac{1}{2} \times 0.25 \times 500). It is an estimate of the consumer surplus to those who would not have made the trip at the higher price.

society, however great his concern for the national

The competition of Rail and Road provides striking

economic solution for the country, so that the head of

Examples of such wrong choices are not lacking.
The competition of Rail and Road provides striking

Examples of such wrong choices are not lacking.

The object of sale at cost is none other than the correct

The object of sale at cost is none other than the correct

The object of sale at cost is none other than the correct

Why Sell at Marginal Cost?

Why Sell at Marginal Cost?

A big manufacturer is considering the installation of
electric furnaces. Depending on whether or not he
carries out his project, it will or will not be necessary
to produce some millions of extra kwh. The issue is
the cost for the economy of these extra kilowatt-
hours, and not the average cost of production per
kwh in the region.

This is particularly clear in France today. Due to
the devaluation of the franc, which has eliminated
a large part of the financial charges of water-power
plants, the average accounting cost of a water-power
kilowatt-hour at the bus-bars of the power stations
is particularly low—let us say 2 centimes per kilowatt-
hour for example. But the cost of kilowatt-hours pro-
duced by new plants, which carry the whole weight
of financial charges on invested capital, is notably
higher—3.50 centimes, for example.

If electric energy is billed, in each region, at the
average accounting price, our manufacturer will not
hesitate for an instant. Even though he may have to
support higher costs on other accounts, he will decide
to establish himself in the water-power zone to take
advantage of the two-centime kilowatt-hour. Con-
sequently the producer must construct a new hydro
plant whose output will cost him, and the country,
3.50 centimes per kilowatt-hour.

At this price, our manufacturer might have given
up his plans; he might have planned to establish
himself in another region where, for a similar price
of electric energy, he could benefit in other respects
from more favorable labor or transport conditions.
In billing energy at the average accounting price
rather than at the cost of production of new plants,
the buyer has been encouraged to make incorrect
choices.

 Doubtless this example is particularly striking be-
cause of the important price difference involved. But
the reasoning remains valid whenever the least gap
appears between average cost and cost of expansion.
Again, with respect to water-power plants, if the
progressive exhaustion of usable sites leads to the
construction of ever more costly plants, the cost of
the energy produced by "marginal" plants is to be
considered and not the average cost of production of
all the plants in service. For it is always the next
plant which is in question.

But, it may be objected, this reasoning is doubtless
valid for new consumers, or for increases in the con-
sumption of old customers: it justifies reference to
the cost of new plants for new consumption, but in
no way prevents the sale of previous supplies at av-

erage accounting cost.

In fact, all consumption is always "new," for the
decision to discontinue it can be made at any moment.
Energy prices ought to be such that the head of the
firm is constantly confronted with the cost which
the entire society will have to bear if he increases
his consumption. But these prices ought likewise to
inform him of the economy which society will be
able to achieve if he reduces his present consump-
tion—either by moving his plants, or by changing
his products, or, most frequently, by improving his
techniques, particularly with respect to efficiency of
use. Now, in an expanding industry, such as ours, the
decrease in the consumption of one customer permits
the satisfaction of the increase in the consumption of
another whose demand would otherwise have re-
quired the construction of a new plant. It is still the
cost of the new plant which is involved, and not the
average cost of plants already in service.

From the limited view of the enterprise itself, the
interest of sale at marginal cost is also clear.

In agreeing to sell a part of its energy produced
continuously at a price below marginal cost, the en-
terprise is heading for a deficit if the consumption of
the beneficiaries of these advantageous prices con-
tinues to expand. And in selling to others for more
What Is the Marginal Cost Involved?

The notion of marginal cost has been introduced previously in contrasting the cost per kilowatt-hour produced by new plants and the average cost of production of all the plants in service. This is a summary view which must be made more precise.

Rigorously defined, marginal cost is the cost of the extra unit.

Costs which are proportional to production immediately come to mind—to the exclusion of all fixed charges—and it may be concluded that this theory is absurd, since it implies that tariffs should provide no payment for fixed charges. And since the theorists of marginalism do not conceal the fact that, in certain cases, sale at marginal cost may cause a loss for the enterprise, it may be thought that the deficit is of the same general magnitude as the total of fixed charges of the enterprise. At this phase of the reasoning, men "with common sense" object, while the "theorists" rack their brains to build up systems which can enable the firm to recover at least a part of this enormous deficit.

It cannot be stated too strongly that this oversimplified interpretation of marginalism is erroneous. If this were the essence of the theory, it would be a matter of urgency to say no more about it.

It is sufficient to recall that, in the same corpus of theory, it can be established that in perfect competition, sale at marginal cost assures the maximization of profit and budgetary equilibrium of the marginal firm. How would this be possible if marginal cost were identical with proportional cost?

Unfortunately, a certain level of abstraction is necessary to untangle the skein of ideas allowing the connection to be made between marginal cost, cost of the additional unit, and marginal cost understood—very generally—in the sense of average cost of production of the additional plant. In default of the ability to offer proofs in this brief discussion, let us take a few examples.

Consider a transmission line for electric energy with respect to which it is assumed:

1. That it is used to transmit a constant load throughout the year (thus we can reason, without ambiguity, in kilowatts).
2. That the costs necessary to compensate for voltage drops are negligible compared to the costs of energy losses.

If the line carries a light load, the additional losses involved in the transmission of an additional kilowatt (throughout the year) are slight. The "marginal cost" of transmission, equal to marginal losses, is therefore very small. When the load under consideration increases, average losses increase, and marginal losses increase a fortiori. For very heavy power loads, average losses are important, and marginal losses are considerable: the marginal cost of transmission is then very high.

To sell transmission service at marginal cost is to bill all kilowatts transported at a price equal to the cost of the losses involved in the transmission of the marginal kilowatt. If the line has too much capacity, its load is light, marginal losses are low, and the sale of the transmission service at marginal cost, although it will recover more than total losses (because average losses are less than marginal losses) cannot recover the fixed charges of the line. If, on the contrary, the line is very inadequate, marginal losses are considerable, marginal cost of transmission is very high, and the sale of transmission at marginal cost insures receipts which are much higher than the total of the cost of losses and of fixed charges.

It therefore appears—and this is an absolutely general conclusion—that sale at marginal cost involves deficits when the firm is over-equipped relative to demand, but it is profitable when the enterprise is very under-equipped.

The conclusion would have been completely different if, falling into the confusion which occurs too often, we had assumed that the marginal cost of transmission was equal to the cost of average losses per kilowatt-hour, so that sale at marginal cost would never yield more than losses—with nothing for fixed charges.

Once this first point is established, it will be noted that the enterprise will not indefinitely choose to remain under-equipped. To avoid the problem of discontinuity, let us assume that the enterprise has a large number of identical lines, providing the same transmission service and carrying the same loads. If these lines are "too" loaded, another will be built. The fixed charges of the whole system will increase by the amount of fixed charges for the additional line, but the total cost of the losses will decrease because the load carried by each line will be less. If the annual gain realized from cutting the losses is greater than the annual cost arising from the increase in fixed charges, the lines had indeed "too" heavy a load and it was correct to build another; a second should also be built if the annual net receipts remain positive, and still another—until there is nothing more to be gained from further construction.

When this level of optimum plant is reached, the
marginal cost of transmission on any line whatsoever of the network will be exactly equal to the average cost of transmission on the last line constructed; for this is only another way of presenting the equality between the fixed charges of the last line constructed and the decrease in losses resulting from its construction.

This is a second conclusion of absolutely general significance: when capacity is optimum, sale at marginal cost of the service rendered by the marginal plant exactly covers the costs of this marginal equipment; so that sale at marginal cost is equivalent to sale at average cost of the marginal plant. The idea of cost of the additional plant, already discussed, makes a reappearance.

It will be noted, moreover, that if all the lines are assumed to be identical budgetary equilibrium for the last line implies budgetary equilibrium for all the lines, and consequently budgetary equilibrium for the entire operation. (This is not, however, an absolutely general conclusion, even though it is approximately true in many cases.) It follows, conversely, that to the extent that the average cost of production of the marginal plant differs from the average cost of production of non-marginal plant, sale at marginal cost may cover more or less than the total costs of the enterprise.

Now we must establish a third point. Let us return to the more realistic case in which (1) transmission service is provided by one line or by a small number of lines; (2) the transmission load increases from year to year in response to the growth of demand. We will continue to assume, however, that (3) the transmission load is constant throughout the year (and that the cost of voltage drops is negligible).

The enterprise cannot adapt its plant capacity each year to the growth of demand. If this growth is such that it equals the capacity of a line in eight years, and if capacity is optimum today, a new line should be constructed in about four years. During the next four years the enterprise will be under-equipped, and the marginal cost of transmission will be higher and higher; then, during the following four years, the enterprise will be over-equipped and the marginal cost, suddenly very low, will progressively increase until in eight years' time it attains its normal level corresponding to optimum capacity.

If transmission tariffs must remain stable—and "tariffs" which do not are inconceivable—it can no longer be a matter of following the true variation in marginal cost from year to year: an average value must be adopted, exactly equal to the value which marginal cost would have if capacity could be continuously increased.

Therefore a third conclusion, also of general applicability (in an expanding industry): sales tariffs based on marginal cost should be established with reference to continuously optimum plant-sizes, regardless of the actual successive phases of over- and under-capacity through which the enterprise passes.

Obviously this does not exclude tariffs equal to marginal costs for the year in question for occasional transmission; but this is not a matter of "tariffs" on whose stability the customer can count.

This third point enables us to reason in the case of optimum equipment in order to establish a fourth and last point.

Let us abandon the hypothesis of a constant load throughout the year: the line transmits a variable load—a "load curve."

At periods when the load is light, the marginal cost is also low and the tariff ought to provide a cheap service. When the load is high, the tariff ought to be high. If the load curve to be transmitted over the line is approximately known—this would be the case of a line serving the consumption of a city—it is possible to associate a different marginal cost with each period of the year and of the day, and to translate these into a sales tariff differentiated into summer and winter, day and night, etc.

And since the line is supposed optimum, the utilization of this tariff will still insure the budgetary equilibrium of the transport enterprise by virtue of the reasoning involved in establishing the second point.

Hence a fourth result of general applicability: the examination of load curves, when these curves have a certain stability, permits the establishment of a marginal cost tariff differentiated according to hours and seasons which represents the cost of an additional unit at different times of the year.

What has been said about transmission lines could be repeated with respect to any other part of the capital plant. The study of "inelastic" installations—that is to say those whose capacity is rigidly determined, unlike a transmission line which can always carry a load heavier or lighter than normal—presents special difficulties which cannot be explored here.

But our object was above all to demonstrate, with an example familiar to those in the electrical industry, the principal aspects of marginal cost tariff-making.

As for principles, the theory seems to us to rest on certain ideas which common sense cannot deny: to sell at cost, and, more exactly, at marginal cost, in such a way that the choices made by the users as to different forms of energy and different ways of using it are guided by the cost of supply for the country.

In its application, the theory requires certain precautions: not to confuse marginal cost and proportional cost; to reason on the basis of optimum capacities; to interpret the case of "inelastic" installations with care.
4.4 Marginal Cost Pricing for Electricity Undertakings

The reason why marginal cost pricing remained removed from practical policy-making for so long is not hard to discern: basically, it appeared that for electricity, for transport, and for other relevant industries and activities, marginal cost pricing would entail a substantial financial loss to the enterprises attempting to apply it. What is the cost of carrying an extra passenger in an empty seat? What is the cost of additional electrical energy, when the system is not using its capacity to the fullest? In both cases the answer is a figure far lower than that which would cover “full cost”; hence the dilemma between marginal cost pricing with deficits on the one hand and budgetary solvency with suboptimal pricing on the other. Probably the greatest contribution of the French economists in this area has been to show that this dilemma can often be more apparent than real, that solvency and optimality need not be inconsistent or contradictory goals.

The problem of electricity pricing can best be approached by assuming a hydroelectric system of the run-of-the-stream type. The output of such a system will be governed by nature and capacity alone—being dependent on the flow of the stream at volumes less than that associated with full-capacity use of the generating equipment, and on the capacity limitation for flows at higher volumes. It will, moreover, entail virtually zero operating costs; and indeed, a casual attempt to apply marginal cost pricing in this situation might well come up with the prescription to price all the electricity produced by the system at (virtually) zero. This prescription is only half correct, however; the proper one is to price electricity at (virtually) zero if the demand for the system’s output is below the level determined by water flow or generating capacity and charge that price necessary to ration the available power among demanders whenever demand at a near-zero price would exceed this level. Since demand for electricity varies widely among the hours of the day, the likelihood emerges that during some hours electricity would be virtually a free good, but during others would command a relatively high price. Moreover, the price would presumably have to vary through the period when it was higher than marginal running cost, because it would have to serve the function of making quantity demanded the same for all such hours. (A run-of-the-stream operation will have significant seasonal variations in water flow, but not within-the-day variations.)

The excess of the price above marginal running costs is in the nature of an economic rent; it is attributable to the scarcity of water in those hours in which the system’s output is limited by the water flow and attributable to the limitation of generating capacity for those periods in which capacity is the factor limiting output. This economic rent can be considered as a component of the short-run marginal cost of providing electricity to any given user since, given that the system is operating at capacity, the only way to satisfy an increase in the demand for electricity by one consumer is to take it away from somebody else, presumably via that small increase in price which would once again limit demand to capacity. And it is in this sense that the charging of a price higher than marginal running cost is consistent with the overall principle of marginal cost pricing.

The above analysis was based on the assumption of given capacity, but the end-product of that analysis—the economic rents generated because of capacity limitations—turn out to be the key element governing decisions about changes in capacity. For example, an addition to the generating capacity of an existing run-of-the-stream hydro system will add to electricity output only in those hours of the year when, in the absence of this addition, capacity would have been the limiting factor. The value of each kwh of output thus attributable to the new capacity will be the price corresponding to the hour in which it is generated; and the income flow assignable as a return on the investment in added capacity will be the excess of the price over marginal running cost—that is, precisely the economic rent referred to in the preceding paragraph. Depending on the size of the economic rent embodied in the price of electricity for the relevant hours of the year, investment in an expansion of capacity will prove to be warranted (for a given criterion rate of return on capital) or not.

In practice, it is obviously impossible to have electricity prices that vary with each hour in the year, so as to perform fully the function of rationing the available capacity among demanders at all times. However, it is possible to approximate this objective by setting different prices for “peak” and “off-peak” use of electricity. The “peak” is determined in large measure by the demand characteristics of the system—being concentrated in the daytime hours on weekdays when industrial demand is dominant, and in the evening hours when residential demand is dominant. Let us assume that a system is made up of identical thermal plants, and facing a growing overall demand for electricity, with a peak of, say,
2,500 hours per year, concentrated in the daytime on weekdays. If we are operating with a given capacity, our objective would be to charge marginal running cost for electricity in the off-peak hours and to set a surcharge for daytime use of electricity sufficiently high so as just to contain demand within the capacity limits of the system. But if we are operating in a dynamic context, being able to vary capacity via investment, we need not worry directly about the rationing of available capacity, but instead should set the peak-time surcharge so as to yield the required rate of return on new additions to capacity. Thus, if new capacity cost were $250 per kilowatt installed and the required rate of return were 7 percent, plus 3 percent for depreciation, we would want to get an economic rent of $25 per year per kilowatt installed, which could be achieved via a one cent per kwh surcharge on all kwh sold during the 2,500 hours of peak-time operation in the year. If capacity were regularly expanded to meet anticipated demand under this rule, we would be (a) charging marginal running cost at off-peak times, and in this sense pricing at marginal cost; (b) charging marginal running cost plus an economic rent sufficient to hold demand down to system capacity at peak times; and (c) earning the required rate of return on invested capital.

The secular increase in the demand for electricity, together with the gradual reduction in the number of unexploited hydroelectric sites and the increasing tendency for electricity generating systems to be combined into large grids or networks, has produced a situation in which thermal costs are the dominant factor governing optimal pricing decisions. Let us suppose, in the example above, that marginal running costs in the thermal plants were one cent per kwh. Then the optimal prices would be two cents per kwh for the 2,500 "peak" hours, and $0.01 per kwh for the 6,260 "off-peak" hours in the year. In terms of those prices, the worthwhileness of any given hydro project can be judged. A run-of-the-stream project costing $1,000 per kilowatt of installed capacity would have a potential maximum yield of $62.60 per year for "off-peak" hours plus $50.00 per year for "peak" hours, and an actual yield quite a bit less than that because of the periods when the rate of water flow rather than installed capacity is the limitational factor. Assuming running costs to be negligible, and using 7 percent as the required rate of return plus 2 percent for depreciation in this case, the worthwhileness of the project would depend upon whether the average expected degree of utilization of the generating capacity was greater or less than 80 percent ($112.60 x 0.8 = $90.08). This type of calculation, based on thermal costs, continues to be valid as long as some thermal plants remain in use at all times. If sufficient run-of-the-stream capacity were available to displace totally thermal operations for, say, the 2,000 hours of lowest system demand during the year, then the optimal pricing policy would be to charge (virtually) zero for power sold during these 2,000 hours, to charge $0.02 per kwh for power sold during the 2,500 hours of peak-time operation, and to charge $0.01 (the marginal running cost of thermal plants) during the remaining 4,260 hours of the year. As long as all thermal plants are identical, and as long as some expansion of thermal capacity is always needed to cover at least part of the secular increase in demand, no more than these three rates will ever be called for in an optimal pricing scheme.

4.5 Electricity Tariff Policy

Since economic analysis in public enterprises has still to gain a foothold in many countries, it might be worthwhile to begin by showing why such analysis is needed. This we do by considering the shortcomings of the very widely held view that electricity tariffs should be determined purely by accounting criteria. These criteria—if only because they are accepted by financiers—have often served the important function of enabling enterprises, in low-income as well as in high-income countries, to mobilize resources to finance much needed expansion. Our arguments, therefore, are mostly about improvements. The high costs of generating electrical energy, and the very large expansion programs in all countries, undoubtedly make striving for improvements worthwhile.

Traditional Approach to Pricing

Unfortunately, a narrow concentration on accounting data—which are relevant to resource allocation only as a poor substitute for unobtainable engineering cost estimates—still typifies some of the tariff advice given to countries that are sufficiently poor to worry about efficient resource allocation. It may be useful, therefore, to describe the typical traditional accounting approach and highlight its shortcomings.

This traditional approach usually begins with a comprehensive stocktaking and evaluation of all assets, old and new, from which, by the application of certain depreciation rules, the annual "capacity-related" or "kilowatt-related" costs are derived. Then there is an evaluation of various running, fuel, and other "energy-related" or "kilowatt-hour-related" costs.

Some costs, such as those for maintenance, have fixed and variable components and are allocated accordingly to capacity- and energy-related costs, respectively. Finally, there are some costs, such as those for metering and billing, that are "customer-related" and not correlated with either capacity or energy demands. The procedure then is to allocate these costs as "equitably" as possible among consumers through the tariff structure—where the notion of equity is that consumers are responsible for covering those accounting costs that they are considered to have imposed on the enterprise. With research into consumer demand patterns (load research), the more advanced enterprises are able to find out how much each consumer class is contributing to the peak and, thus, to the capacity-related accounting costs. Then energy- and customer-related costs are added in and a "cost-based" tariff is formulated for each consumer class. Typically, a consumer may have one, two, or some combination of three elements in his bill: a fixed or minimum charge (to recover customer cost); a kilowatt charge related to his contribution to capacity cost; and a kilowatt-hour charge. Simplifications are often sought, some resting on complicated analysis which nevertheless leads to a simpler tariff and which meets the general aims of the traditional approach. For example, frequently it is found that consumers who take more kilowatt-hours take relatively less kilowatts at the time of peak demand. When peak kilowatt and kilowatt-hour consumption are empirically related in this way it is possible to simplify the tariffs, eliminating the kilowatt charges by incorporating them in the kilowatt-hour charges. Consumers (even very large consumers) are then given declining-block kilowatt-hour tariffs, because their total capacity requirements for each kilowatt-hour consumed decrease with the number of the kilowatt-hours consumed. In addition, the fixed or minimum charge may be added onto the first block, leading to a very simple tariff related only to kilowatt-hour consumption. Provided the empirical relationships used to derive this simplified tariff hold, the simplified tariff will certainly meet the equity principle that customers should pay for the accounting costs allocated to them.

This is, albeit in simplified form, the basis of an ideal accounting approach. An enormous amount of information is collected and manipulated in the process, and enterprises paying consultants to do this sort of thing find it a very expensive and time-consuming exercise. Diversions from these ideal tariffs—sometimes very substantial diversions—often creep in, but for now consider the ideal accounting tariff in general terms.

The first limitation of this approach is that, except by chance, prorated accounting costs are quite different from the costs relevant to resource allocation. One reason for this is that the accountant is concerned with recovering sunk costs whereas, for efficient resource allocation, it is the actual resources used or saved by consumer decisions that are important. Prices are the amounts paid for extra consumption, and need to be related to the incremental costs in meeting extra consumption. If new consumers are connected to the system, or if existing consumers increase their consumption (for example, during the system's peak), additions to generating and network capacity may be required. It is important, therefore, that prices should signal to consumers the costs of such consumption changes. The argument works the other way, too. If consumers reduce consumption (for example, during the system's peak), such costs are avoided, and if prices reflect these costs, the savings on their bills will equal the resource savings. Hence, prices should be related to the value of resources used or saved, and the valuation of these resources—the estimation of costs—requires a forward-looking estimate. The backward-looking estimate of the traditional approach creates the illusion that resources which can be used or saved are as cheap or as expensive as in the past; that is, that resources are as abundant or as restricted as in the past. On the one hand this may cause overinvestment and waste; on the other, it may lead to underinvestment and unnecessary scarcity. In addition, if the past holds a number of poor projects, the sunk costs of mistakes, if reflected in prices, will overstate the costs to the consumer of extra consumption, which is not efficient.

Another reason why prorated accounting costs differ from those relevant to resource allocation is that the tariff schedules and the various simplifications thereof are derived by spreading total accounting costs among consumers. Broadly speaking, this generates tariffs which relate to average rather than to marginal costs. But for efficient resource allocation, prices should be related to the resource costs of changes in consumption; that is, what is needed is pricing according to marginal, not average, cost. The addition of a new consumer or an increase in the consumption of an existing consumer will impose additional costs on the enterprise, while a reduction in consumption will save costs. These alterations in costs are the ones that need to be reflected in tariffs. The change in the cost to a consumer of altering his electrical behavior will then mirror the change in the cost to the enterprise.

This brings us to the second limitation of the accounting approach. Fairness or equity in the approach is couched in rather narrow terms: consumers should pay for their allocated share of accounting costs. These allocated costs, as previously explained, may very well differ from the costs that consumers are causing the enterprise to incur. Apart from this
fact, it is evident that such allocation of costs involves judgments that may be arbitrary. Fairness is surely an attribute of tariffs considered in relation to consumers, not of costs considered in isolation. As such, it depends on, for example, whether a consumer is rich or poor, and whether he deserves special concessions. As we shall make clear later in this chapter, we fully accept the notion that questions of equity are relevant in tariff-making. Because of the huge capital requirements of the power sector, we also accept the point that electricity tariffs should often yield revenues sufficient to cover accounting costs and, in addition, should make a substantial contribution to the self-financing of future system growth. Electrification projects in backward areas are possible exceptions. But fairness, revenue requirements, and cost analysis require separate analyses in tariff making. First we must analyze costs, then consider revenue requirements and bring in views of what is fair. Compromise may be necessary, but it is absolutely necessary to start off with a purely objective analysis of costs.

The third limitation of the accounting approach stems from its neglect of the incentive effects of tariffs. Even if the allocated accounting costs were equal to the marginal costs of resources, and equity issues were unimportant, this still would be a serious defect. Tariffs often, if not always, have to be simpler than the cost structure they represent. Generally, billing can be done only monthly at most; there are restrictions on how much prices can be varied in response to random changes in demand and supply conditions; and elaborate metering both is too costly for all but the largest consumers and, because it bewilders many consumers, can be counterproductive. Simplification is a central part of the formulation of policy on tariff and metering. But how do we simplify without nullifying the aims of the tariffs? We have found that the answer, perhaps not surprisingly, depends on the aims of the tariffs—simplified tariffs designed with only accounting aims in mind may differ enormously from those suggested by economic analysis. It is easier to show this through a concrete example.

Consider, for example, the problem of charging a consumer for capacity costs induced by his demand at the time of the system's peak. The traditional approach may use any one of several devices to make sure that the consumer pays for the costs incurred. There may be a fixed charge related to the consumer's demand at the time of the system's peak demand, or, as discussed above, capacity charges may be incorporated into the energy charges, based on an empirical (but not a causal) relation between consumers' energy consumptions and demands for capacity; or there may be an empirical relation between the size of a consumer's house (measured in terms of floor area) and his demands for capacity, so that capacity charges may be related to floor area. These are only three examples. Very simple single or two-part tariffs, with the "variable" part having a declining-block pattern, can be developed on such bases. It is evident that, while such devices may satisfy the accountant's equity principles, only a charge related to consumption during peak is likely to provide the consumer with the incentive to economize—and this is, of course, on the proviso that he knows what he is being charged for and that he knows approximately when the peak is occurring.

From an economic viewpoint, tariff simplifications should be designed to retain the incentive effects as far as possible. The entire design of the tariff and metering policy turns on this aim. As far as is possible, consumers should know when consumption is expensive and when it is cheap. Declining-block tariffs, however ingeniously concocted, do not tell consumers that peak-hour consumption is expensive, nor do kilowatt charges that are related to, for example, the floor area of the consumer's house. Time-of-day metering will; so will peak-load limiters if they are properly adjusted; so, under certain circumstances, will kilowatt charges related to a consumer's demand during peak. The essential condition which declining-block and floor area tariffs fail to meet is that of making peak-period consumption more expensive than off-peak consumption.

An entirely different tariff and metering policy may follow when incentive effects of tariffs on consumers' demands are considered. We have found that, in most countries, financial questions and accounting rules dominate the level, structure, and types of tariffs. Incentive effects are sometimes considered, but generally only as an afterthought or out of the necessity of holding back a very rapidly growing peak demand (often 20 percent a year or more in developing countries). What is needed, in our view, is the opposite philosophy. Incentive effects should be considered first when choosing tariff types and meters. Financial targets can still be set and deserving consumers can be given concessions, but this should be done in a way that has the least damaging effect on incentives.

We can now state our approach in general terms. We accept the importance of financial targets and we accept the importance of equity—although we disagree with the traditional accounting approach to the meaning of equity. We also stress the importance of resource allocation. We begin with resource allocation by analyzing cost structure and the incentive effects of the various simplified tariff and metering policies open to the enterprise. Next, if required, come questions of equity to allow, for example, for the problems of low-income groups. Finally, financial targets enter into the picture as we confine any necessary increases to certain elements of the tariffs so
that the least possible damage is done to equity and resource allocation.

From what we have just said, it is natural that we proceed in sequence to consider costs for pricing, tariff simplification and incentives, approaches to cost distortions (shadow prices, second-best), and equity and finance. There is, however, one technical matter to clear up first, and this is the interaction of pricing and investment.

Interaction of Pricing and Investment

The approach to investment planning usually begins with a forecast of demand, and this is followed by a search for a least-cost policy for investment using one or another of the various cost-minimizing techniques. Once the least-cost policy for investment has been determined, the cost structure relevant for pricing can be derived (sometimes directly from the output of the cost-minimizing technique). At this stage it might be argued that new prices set on the basis of these costs would alter the demand forecasts so that the investment program should be revised accordingly, giving a new cost structure. But this iterative procedure may be an unnecessary refinement. In practice, prices can only be adjusted slowly, often with a considerable time lag for debate and approval. Prices also take a long time to act, since the demand for electricity is linked largely to the stock of electrical appliances and machines. Anyway, by the time that price revisions have been made and have begun to have their effects, the time will have arrived for revised forecasts and programs. The effects of prices on demand should be evident in the trends in the revised forecasts. Thus, while the feedback of prices on forecasts and the investment program is important, feedback is best dealt with iteratively (as in automatic control systems) by waiting for prices to begin to have their effects on forecast demand levels, and changing investment plans accordingly.

This process of gradual adjustment and of slow response to prices is fortunate in that we need not bother with the value of price elasticities, which are notoriously difficult to estimate reliably. On the other hand, the process is unfortunate if prices are badly out of line with costs, for then it may take a long time to put things right. This raises a dilemma for project justification. A project may show poor economic returns on current prices because the prices are too low when the project is operating. But if price adjustments have to be gradual and price response is slow, we still may have to accept the project rather than make matters worse and accept physical rationing. Until price reforms have had their effect, decisionmakers may have to accept projects with apparently poor economic rates of return, however unwelcome they may find such a situation.

Costs for Pricing

We begin with a theoretical issue—pricing according to short-run versus long-run marginal costs. It is now many years since their equivalence under conditions of certainty was demonstrated by Boiteux and other French authors. The important point is that there are definite limitations on the possibilities of using prices to ration available capacity when demand and supply conditions turn out to be different from those expected when prices were set. Thursday's prices cannot be raised because some turbines do not start on that morning, and January's prices cannot be raised if it becomes apparent that the month is colder than usual—if only because meters are read just once a month or because consumers get irritated with sudden price increases. Consequently, physical rationing (load shedding) is unavoidable at times. In the cost function we include a term representing the expected social costs of load shedding, including losses of industrial output, and the nuisance and cost to consumers of having to substitute candles, batteries, or oil heaters, or of having to do without. Thus an increase in the level of (the probability distribution of) demand in the short run—that is to say, when capacity adjustments are not possible—will increase both the expected demand that is met and the expected demand that is not met. The expected costs of outages and expected fuel and other variable costs also rise. Thus the relevant cost for short-run pricing is a probability-weighted average of the marginal costs of not meeting demand and the marginal variable costs of meeting it. (Capacity costs do not change in the short run.)

In the long run, capacity adjustments can be made to keep the probability of interruptions down to an acceptable level. The long-run marginal cost of supply is the marginal cost of extra capacity plus the expected costs of extra output. (There is also a rule to define the optimal level of extra capacity: the marginal costs of extra capacity should equal the expected social savings from marginal reductions of supply interruptions.)

Once short-run conditions become different from those anticipated when long-run capacity decisions were made, one might as well optimize for the short run. This point derives its force from the length of time it takes to plan and install new generating plants—six years is typical. If demand has grown faster than anticipated, it will take a long time before the rate of growth of capacity can be accelerated. Meanwhile, it may be preferable to raise the price of electricity during periods of peak demand rather than to accept an increased probability of power cuts.

Accepting this point in principle, we have found that there are situations where it is not very helpful.
The argument assumes that the tariff structure is such that there are appropriate peak kilowatt-hour prices or kilowatt charges that can be raised. But this is not always the case. The tariff structure may contain no such elements. The problem is then to suggest a new structure, not to alter some elements of existing tariffs. In addition, information generally is not available on the expected costs of load shedding, and this, too, adds to the difficulties of applying the short-run rule.

Simplifications and Incentives

The long-run marginal costs of supply can be stated very simply for most systems, though in practice a lot of work is needed to estimate them for the various voltage levels of service. Most writers derive rules about marginal cost for public enterprises having nonstorable inputs (or where storage is without cost), so these rules apply only to predominantly thermal systems. They have two elements: marginal energy (fuel) and running costs: plus (during peak times) marginal capacity costs. We have adjusted these rules to allow for indivisibilities in capacity expansion. We show that it is best to consider the marginal capacity costs as being the present worth of the costs of bringing forward capacity expansion by $T$ years, averaged over $T$, where $T$ is the interval between investments. In many cases, this is approximately the average of incremental costs of expansion. Additionally we have derived the rules for mixed hydro-thermal systems, that is, where storage is important. In the simpler cases, marginal energy cost is zero in the wet or filling-up seasons. In the dry or discharge seasons, however, marginal energy cost rises to: the cost of adding to storage capacity (long-run rule); or the rationing price needed to keep the dry season energy demand down to the point where the probability of interruptions is at an acceptable level (short-run rule); or the fuel cost of thermal plant, if this is used to provide for extra output, less net capacity savings (over hydro) if extra thermal plant is needed (again, this is a long-run rule).

If marginal cost can be described so briefly, why is it so difficult, technically speaking, to reflect accurately in tariffs? We have touched on the answers earlier: high variability and unpredictability in demand, and cost and bother in metering. For many large industrial, agricultural, and commercial consumers, even quite sophisticated metering may cost less than 0.1 percent of their bills, so only the first of these problems is important. Meter reading can be done monthly, so seasonal variations in prices are possible. The main problem lies in determining first the times of day when demand is pressing on available capacity and then a means to meter and charge for it. There is a fair range of choice. Supplies that are subject to interruption can be offered at a lower price to those consumers for whom occasional interruptions are not too troublesome. Other consumers may be willing to pay for the high costs of supply during peak; for these, we may use either maximum demand meters, switched on during the peak hours of the month, with a charge related to maximum recorded demand (in this system, kilowatt-hours are metered separately and related to fuel and running costs of supply), or time-of-day meters with high kilowatt-hour rates during peak hours. Telecontrol can be used to make timing of the metering more flexible and accurate. Broadly, we have found that the choice between metering maximum demand and metering kilowatt-hour consumption during peak depends on the demand conditions. If the peak demands are persistent over many hours, in the form of plateaus, it is worth casting the net widely, so to speak, by applying uniform kilowatt-hour charges over all peak kilowatt-hours during the month; off-peak kilowatt-hours are recorded separately. If, however, the peak demand is spiky, a wide net is too restrictive while a narrow one may miss the peak or cause it to move elsewhere; here, a charge related to observed maximum kilowatt demand during potential peak hours of the month, with a separate meter for kilowatt-hour consumption, is more appropriate.

Turning to small, mainly domestic, consumers, metering and billing costs and the lesser sophistication of the consumer necessitate simple tariffs. (A time-of-day meter, for example, may cost about $50.00 and is only worthwhile for domestic consumption at levels typical of Europe and North America.) Tariffs for small consumers can reflect only one or two of the various features of the cost structure. As there are quite a few features in the cost structure, this raises the problem of choosing which to reflect and which to average out. Briefly, seasonal tariffs can be applied even if billing is monthly or bimonthly; less frequent billing may preclude their application unless there are only two seasons (wet and dry) within the year. The problem arises with charging for peak-off-peak differentials within the day. At the simplest level, a flat kilowatt-hour rate may be charged, pitched somewhat below the marginal costs of peak demand but somewhat above the marginal costs of off-peak demand to avoid undue discouragement of off-peak consumption or undue encouragement of peak consumption. At a slightly more advanced level, a load limiter tariff can be introduced, where consumers subscribe to a certain maximum demand and are automatically and temporarily disconnected by a small circuit breaker (in the house) if they exceed it; in addition, consumers pay a flat kilowatt-hour rate. Finally, for the larger domestic consumers, it may be worthwhile to have a time-of-day tariff.

This does not exhaust the options open for eco-
nomic tariff and metering policies—in particular, telecontrol may offer useful new options—but at the current state of the art, these seem to us among the most important ones (at least in developing countries).

There are also some common institutional problems which can have an important bearing on the choice of tariffs and meters. For example, monthly meter reading is labor-intensive, which increases costs. One might expect it to be more useful in developing countries where labor is cheap. But such countries often find great difficulty in recruiting and managing reliable meter readers, and the main reason for monthly billing seems to be that consumers default less on twelve monthly bills than on fewer, necessarily larger bills.

Reliability, whether of meters or of meter readers, is not only a matter of training and maintenance effort. Consumers can cheat, either by tampering with their meters or by suborning the meter readers. Indeed, consumers in some countries are adept at stealing electricity, at some personal risk. Furthermore, billing systems can degenerate so that bills arrive late, are inaccurate, or fail to arrive at all. Collection may go wrong, too.

Finally, it is evident that consumers must have an understanding of what it is they are being charged for. The purpose of a tariff structure is to provide consumers with incentives. A high rate provides both a message that special economy is called for and an encouragement to do something about it. But consumers by themselves may not understand the message and may not know how to do anything about it. They need help, and it is part of the job of the electricity enterprise to provide it. Tariff making very much needs to be supplemented by technical advice. For large consumers, the advice of commercial engineers is needed. For small consumers, individual advice is not practicable, and, though tariffs are already simple, a certain amount of advertising still may be required. By helping consumers adapt to the tariff structure, electricity companies help themselves.

Cost Distortions

A tariff structure which fully reflects the enterprise's cost structure may not lead to efficient resource allocation. One reason relates to what is known to economists as the "second-best" problem. Consumers' choices may be influenced not only by the price structure of electricity but also by the prices of other fuels. Again, the prices paid by the electricity industry may not reflect the value to the economy of the resources used. Only if all such prices are right in some sense have we the "first-best" situation where tariffs that reflect cost will lead to an efficient allocation of resources.

One answer to the second-best problem, which may be institutionally necessary although intellectually unsatisfying, is to ignore it. A practical reason for this is simply that the problem may not be in one's terms of reference. The electricity enterprise has been given the job of supplying electricity, not of running the whole economy, and may have to work within the existing framework. If the enterprise thinks that the tax on oil makes oil prices too high, for example, it may nevertheless have to mind its own business. (Another practical reason for ignoring the problem is that distortions elsewhere in the economy should be tackled directly.)

The other answer is, of course, to try to make some adjustments for the distortions. Such adjustments may be called for not only in setting tariffs but also in making investment decisions. It is in this latter context that much has been written about the use of shadow prices as a way of making the adjustments in developing countries. Government directives are sometimes provided to public electricity companies to use appropriate shadow prices for capital, labor, foreign exchange, and material inputs when comparing projects: this, at least, facilitates the institutional problems (not to be underestimated) of deciding on an efficient investment program.

Given that shadow prices are important for determining which projects make up an investment program based on least-cost, conflicts may arise when attempting to reflect them in tariffs. In power systems with fuel oil plants, for example, taxes on fuel oil, which are generally heavy, may rightly be neglected when power plants fired by fuel oil are being compared with, say, plants driven by hydro power, since taxes are not a cost but rather the government's share in profits. However, the government will still expect its taxes to be paid, so the enterprise must somehow raise them. This has three implications for tariffs: their structure should reflect resource cost (computed using shadow prices) as far as possible; the revenue which this structure would generate should be computed; and then, if this is too low, the structure should be amended to achieve the desired financial objective by increasing some elements of the tariffs in such a way as to do least damage to the efficiency and equity objectives.

Before passing to the equity and financial aspects of tariffs, there is one final point to make about distortions to costs. Tariffs which reflect resource costs would often lead to considerable financial surpluses in the electricity enterprises in developing countries (fuel oil taxes notwithstanding) because these enterprises are among the heaviest and most intensive users of capital and foreign exchange, both of which may have high shadow prices. These en-
Enterprises generally have had small beginnings and have faced very high growth rates of demand, and thus, large expansion programs over long periods, so it is common for them to experience considerable difficulties in meeting financial targets. Thus, strong enterprises generally have had small beginnings and shadow prices will not always be accepted.

Finance and Equity

We have already noted that considerations about equity and finance as well as about resource allocation are relevant to tariffs. Our view is that it is best to design a tariff that reflects cost, subject to any second-best considerations, and then to modify it if the promised revenue is inadequate or if it seems unfair. This procedure, which keeps resource allocation, equity, and finance separate in the earlier stages, avoids the confusion that afflicts the traditional approach.

We have only two general points to make on the financial side. First, the management of an electricity enterprise often welcomes a requirement that the enterprise earn a certain minimum rate of return, calculated so as to cover its accounting costs and possibly to earn sufficient revenue to finance a certain proportion of future capital expenditure. This requirement helps support financial responsibility, mobilize financial resources for expansion, and may enable management to obtain considerable autonomy in running the enterprise, a valuable spur to innovation and efficiency. Enterprises which rely on having their deficits met by government are usually (though not inevitably) in difficulties, and their capacity to innovate is sometimes noticeably stymied. It is true, we admit, that very high profits may have a debilitating effect, too, but this is a much less common phenomenon.

Our second point is that putting tariffs at a level above that called for by pure cost reflection is equivalent to taxing the supply of electricity. The yield of the "tax" may finance the enterprise's capital expenditure rather than flow into the country's exchequer. Apart from any effect on the enterprise's independence of the sort just noted, if the exchequer would otherwise finance part of the enterprise's capital expenditure, the difference is only one of bookkeeping. What concerns us here is a point relating to second-best resource allocation. Something does have to be taxed and there is no general case against taxing electricity. Indeed, if other kinds of energy are taxed, there is a positive case for taxing electricity as well, so that relative prices including tax reflect relative cost before tax. More generally, in countries where the tax base is limited and the tax system is inefficient, electricity is a good candidate for taxation even if the tax takes the implicit form of a high proportion of self-finance. Although there may be substantial institutional difficulties in raising financial targets, therefore, it is something to be encouraged and is worth striving for.

Fairness in tariff structure is a more contentious topic. We have already argued that there is nothing intrinsically fair in the financial view that tariffs should be set so that revenues from each class of consumer cover the share of accounting costs allocated to that class. Similarly, there is nothing intrinsically unfair in revenues that do not cover accounting costs—as happens in the village electrification programs of many countries. Certainly a tolerant attitude on tariffs to poorer consumers can be defended in such cases when the more appealing methods of redistribution through fiscal policy are not administratively feasible. Having said this, however, we also state our general belief that distortions in the use of economic resources stemming from inequities in the economic system, like those stemming from distortions in the pricing system, are best dealt with directly. Countering inequities through cross-subsidization within services such as electricity is a poor second best.

Returning to less controversial matters, we have noticed that what interests electricity enterprises and governments is often not the fairness of tariff levels in relation to costs but the political acceptability of tariff changes. Since no one ever objects to a reduction in tariffs, our observation is that if a new tariff would noticeably increase the bills of a group of consumers who would create a big fuss, that tariff will be reexamined. A large part of the effort involved in setting a new tariff is spent comparing it with the existing tariff for a sample of consumers in order to estimate the effect on their bills. Thus, in practice, the "fairness" constraint on new tariffs often becomes an implicit stipulation that no consumer shall suffer an increase in his bill of more than so much percent within a certain period. The economist who is advising on tariffs has to take this kind of constraint into account, even if only by recommending a gradual rather than a sudden transition to a new tariff. The people who run electricity enterprises naturally value a quiet life.

4.6 Energy Pricing Policy

The purpose of management concerned with energy is to reduce the consumption of energy per unit of gross domestic product (GDP), and to control the growing demand for energy so as to reduce its overall...
cost to the economy (for instance, by fostering the use of indigenous energy sources to replace imported energy).

The demand for energy may be influenced in a number of ways, directly or indirectly, which may have either short- or long-term effects. Energy pricing is one of the most effective management tools for controlling demand. Others are: administrative, such as rationing and allocation schemes; legal, such as the prohibition of certain energy-using equipment (private power generators and automobiles of more than a certain horsepower); fiscal, such as the imposition of discriminatory taxes on certain fuels or energy-using equipment. Differential import tariffs, variable electric power tariffs, and sales taxes on gasoline are fiscal measures commonly used to manage energy demand.

In some cases, the initial reason for imposing such taxes was to raise revenue for the government; the aspect of energy demand was neglected until recently. In the case of electricity tariffs, multiple rates are often initiated by utilities to reflect the cost of serving different classes of customers. It is not uncommon, however, to find that the tariff systems of state-owned utilities have been modified to serve social or political ends. One of the commonly used, but least effective, methods of managing the demand for energy is to appeal to the public conscience to conserve energy in one form or another. While this may have an appreciable effect in the short term at a time of national crisis, it becomes less effective each time it is used and cannot be regarded as a substitute for more positive action over the long term.

Short-term measures to manage the demand for energy seek to reduce existing inefficient and wasteful practices, such as using excessive display lighting, travelling alone by car, and altering thermostat settings in buildings. These measures nearly all restrict consumption in some way, and in developing countries often affect only a very small proportion of total energy demand.

Long-term measures seek to reduce the growing demand for energy in relation to gross national product (GNP) by such methods as: replacing existing energy-consuming equipment with new, more energy-efficient types; modifying old buildings and constructing new ones to minimize energy consumption; changing energy consumption patterns from the use of imported fuels to the use of indigenous energy sources; and planning industrial growth in accordance with availability of indigenous energy sources.

To achieve these changes, governments should use the full range of methods available in a coordinated fashion. For example, if import duties are adjusted to favor the import of energy-efficient equipment, pricing policy should be used to reinforce this. Placing a high import duty on vehicles that consume a lot of fuel has little effect if the corresponding fuels are still sold at low, subsidized prices. In a developing country, legal and administrative methods of controlling energy demand are usually difficult if not impossible to enforce, so greater reliance should be placed on fiscal and pricing mechanisms. Although pricing policy is probably the most flexible and effective management tool for controlling the demand for energy, many governments are reluctant to use it vigorously because energy price increases invariably arouse serious social and political opposition.

Energy Pricing Policy

Energy pricing policy is at the core of a country's overall energy policy. It is directed at a series of related issues such as:

(a) What should the relation be between the prices of primary energy fuels such as coal, crude oil, and natural gas charged to power utilities? And what relation should these prices bear to those charged to the producers?

(b) What should the relation be between the prices of petroleum products (liquid petroleum gas, gasolines, kerosenes, diesel, and fuel oils) at the ex-refinery price, as well as at the retail price?

(c) What implications does a specific energy pricing regime have for the future market shares of primary energy fuels and secondary energy supplies (such as electricity) in the total energy supply mix?

(d) At what rate should indigenous nonrenewable energy resources in developing countries be depleted, and does the price applied to any given resource satisfy the objectives of the depletion rate?

(e) What should the relation be between the price of competing fuels at the end-point of consumption, for example, between charcoal, liquid petroleum gas (LPG), kerosene, natural gas and electricity for domestic cooking; between gasoline, diesel oil and LPG as fuels for transport; between coal, fuel oil, and natural gas for thermal power generation?

(f) Should energy prices be used as a mechanism to achieve the objectives of income distribution in a society? If so, how?

The complexity of the above questions is such that answers can only be provided in a country-specific setting. Nevertheless, guidelines can be used to delineate the appropriate range of prices of primary nonrenewable fuel resources, such as coal, crude oil, or natural gas.

(a) The delivered price should not be higher than the economic costs of the next best alternative fuel delivered to the particular intermediate
and maintenance costs of equipment, required to utilize and the associated differences in capital, as well as operation to account fully for quality differences of alternative fuels.

The issue of reflecting future scarcity of particular fuels in present prices for these fuels becomes significant. This is particularly so in those developing countries where the indigenous energy resource base of particular fuels is very limited in relation to present and future demand, and the country faces, upon depletion of a given resource, the need to replace it by purchasing energy in the international market. The problem is even greater when the international availability of fuels is constrained and prices are higher than those previously charged for the domestic resource.

Some developing countries have taken positive steps since 1973 to adjust their liquid fuel prices to international levels. However, oil-exporting developing countries (both OPEC and non-OPEC members) and some oil-producing but nonexporting countries still tend to sell petroleum products in their domestic markets at prices well below international levels.

Direct subsidies applied to specific liquid fuels—no matter how justified socially—severely distort demand patterns for petroleum products and give rise to abuses as well as interfuel substitution, since many items of energy-consuming equipment can easily switch fuels, especially if the financial incentive to do so is high. For example, with slight modifications a gasoline-kerosene admixture can be used in motor cars, kerosene can be mixed with diesel fuel, and with adjustments, LPG can substitute 100 percent for gasoline in transport use. When the retail prices of kerosene and gasoline are permitted to differ by as much as 300 to 400 percent, it is inevitable that kerosene will be diverted to uses other than those originally intended to be subsidized.

Energy pricing can play a considerable role in accelerating exploration for and development of primary fuels. Several energy-producing entities in developing countries are state owned and are unsound financially because of government pricing policies. These policies do not permit state-owned energy producers to charge tariffs or prices sufficiently high to enable them to recover the cost of providing services to their customers, or to accumulate sufficient financial surpluses to pay for expansion of and exploration for new resources. The results are poor service to the consumer and, ultimately, supply shortages.

4.7 Cost of Using Roads

What does it cost to run a truckload of farm produce to the market? To be more specific, what are the costs of running the loaded truck twenty miles over a country road, then two miles through suburbs and into the center of a large town? To the farmer, this seems a simple enough question: there is the cost of gasoline and oil, the cost of the driver's time, and the cost of maintaining the truck. Even to the road engineer, the costs of road use may seem simple: so much for maintenance a year, with a simple rule of thumb for additional costs for the wear and tear of the heavier traffic in the suburbs and the town center.

When questions arise about widening and regrading, or of replacing the old road by another, economists and planners are faced with problems whose intricacy is equaled only by their interest and importance for economic development. Is the cost of the investment to be paid for by the users, or should the community as a whole pay?

Appropriate Concept of Cost

The definition and measurement of the appropriate concept of cost is crucial to any study of road use charges. Any particular concept of cost is uniquely related to a particular and well-defined decision: the amount of use made of the roads is determined by individuals deciding whether or not to make a journey. Each person takes into account the conse-


7. Both economic and financial costs should be adjusted to account fully for quality differences of alternative fuels and the associated differences in capital, as well as operation and maintenance costs of equipment, required to utilize these alternative fuels.
quences of making a journey (or perhaps another journey); that is, the cost of a vehicle journey is measured in terms of the resources (i.e., the goods and services) forgone in order to undertake that journey.

The costs of the farmer going to market—fuel, wear and tear, and driver’s wages (or his own time)—appear in his accounts and he is legally obliged to meet them. They constitute the private cost of the journey.

The government road authority will pay for the wear and tear of the surface of the highway. This is called the variable maintenance cost and, assuming no user taxes, will not appear on the books of the motorist. To distinguish it from the private costs borne by the motorist it will be called a part of the social cost of the vehicle journey.

When traffic is so light in relation to the capacity of the highway, that vehicles in no way impede one another’s speed and progress, the sum of the private costs and the social costs (repairing the damage done to the surface) represent in full the costs of the vehicle journey. It is the sum of the private cost and the cost of the damage done to the surface of the highway.

It is important to notice that this cost does not include the costs of providing the highway; nor does it include those costs of maintenance which are unrelated to the vehicle journey (the invariate maintenance costs is the term used to describe them). The resources invested in the highway were committed in the past; they were sunk and the costs are bygones. No fraction of these resources could be saved if the motorist forbears to take his journey; consequently, they are not part of the journey costs.

Up to this point we have supposed that traffic was so light that no vehicle impeded any other; each driver could choose whatever speed he liked. This assumption is approximately correct for most rural, farm-to-market, and interurban highways. But it is clearly incorrect for urban streets, for most of suburbia, and for some interurban highways; in the example given above, it would probably be correct for the first twenty miles but incorrect for the final two miles. And the center of a large market town is only one example of a place where vehicles do get in one another’s way and there is much congestion.

Under congested conditions an additional journey will add to the congestion. The vehicle will get in the way of other vehicles using the road and will cause their costs to increase as they waste more time in traffic jams and incur higher maintenance costs per mile in the dense traffic. Thus the decision by a vehicle owner to use a congested highway involves all other users in increased operating costs. But these additional costs appear as spread among all other motorists using the highway. They are not the responsibility of any particular motorist who decides to use his vehicle on the congested road; they do not appear in his accounts. On the other hand, these “congestion costs” are clearly attributable to the vehicle journey; if the vehicle journey had not taken place they would not have been incurred.

To sum up, then, the costs of a vehicle journey consist of:

- **PRIVATE COST**
  - (a) operating cost—borne by the motorist himself;
  - (b) variable maintenance cost—borne by the public road authority;
  - (c) congestion cost—borne by all other users of the road.

The central problem of a policy for user charges is to deal with (b) and particularly (c).

**Reflecting Costs in Motorist’s Accounts**

Since the sum of these three elements constitutes the resources (i.e., the goods and services) forgone due to the vehicle journey, it is clear that all three costs should be reflected in the motorist’s accounts. If these costs are not properly exacted from the motorist, he will be induced to undertake too many journeys, and thus to add unduly to the congestion. High traffic congestion is typical of large conurbations throughout most of the world, but user charges rarely reflect adequately the cost of such congestion, and traffic jams and snail-like speeds are the consequence. These are the wastes of user charges that are too low.

If, on the other hand, user taxes exceed the sum of the variable maintenance cost and the congestion cost, then the vehicle owner will be dissuaded from undertaking certain vehicle journeys although the true cost is less than the returns. Taking into account the user tax, he may find that the sum of the operating cost and user tax exceeds the returns he expects from the trip. The journey is actually worthwhile but the unduly high tax prevents him from making it. Consequently, potentially valuable services of the road are wasted. Such conditions are probably typical of most interurban roads and rural.
highways. The waste is less obvious than in the example of the congested cities—but it is there and just as pervasive.

The importance of linking costs to the decision about road use can be illustrated by considering what is probably the most common mistake with the cost approach: surely, one may argue, the additional vehicle journey causes the road authority to undertake new investment in the highway, and the costs of the new investment are therefore part of the costs of the vehicle journey. The fallacy in this chain of reasoning is the allegation that the vehicle journey "causes" an increase in road investment. It doesn't. The government decides on the level and distribution of road investment. Even though traffic has increased, it might decide to spend the same amount, it might decide to invest more, or it might even reduce investment in roads. All these decisions may be perfectly rational. Decisions about expanding and improving the road are quite different from decisions about whether or not to use the existing highway. To confound them merely spreads confusion.

What is worse, the wrong analysis leads to uneconomic user charges and uneconomic use of the available road space. If user charges are fixed according to the measure of the investment cost "caused" by an additional vehicle, one can easily see that the charges will be too high on the uncongested interurban and rural highways, and will probably be too low on the urban streets. Consequently, the urban highways will be overcongested and the rural and interurban roads will be underutilized. These wasteful effects of such definitions of costs, and of the derived system of user charges, are the really telling indictment against them. User charges are based on "cost," but the "cost" concepts are wrong; they do not pertain to the decision of the motorist whether or not he will use his motor vehicle for a journey along the road. The relevant costs are those that are incurred due to his particular use of the road.

But, it might also be argued that the additional vehicle journey that the motorist undertakes will play a part in inducing the authorities to build bigger and better roads. An increased investment in the highway is then a consequence of his vehicle journey, and so the cost of the investment should be reflected in the user charge. This line of argument provides a rationalization for making the user charge equal to the cost of providing additional road space—in other words, fixing the price for using the road at the long-run marginal cost of supplying new road space.

But the long-run marginal cost is useful as a criterion for user charges only when it is equal to short-run marginal cost. When the long-run cost differs from the short-run cost, it is the wrong basis for pricing policy and will cause too much—or too little—use of the road. The short-run marginal cost is always the appropriate value at which to fix the user charge.

Since the unit of output may be chosen as large or as small as one likes, it is often thought that the concept of marginal cost is arbitrary. This is, however, not correct. The relevant unit is the one that faces the user. It is both physically possible and administratively efficient for road users to be permitted to purchase any quantity, however small, of road service. The user takes into account the price of the little more or the little less in making his decisions. Thus the cost concept should measure the cost of the little more or the little less.

**Investment**

This structure of user charges, along with all others, has some disadvantages. The most critical is the absence of any simple easy-to-observe guide to the appropriate amount and distribution of investment in highways. The common-sense criterion is: "Build a better road if the reductions in vehicle operating costs and in the annual maintenance costs of the highway are at least as large as the annual cost of the road." But this applies only to existing traffic; it cannot be applied to new traffic generated by the reduction of transport cost. The development effects of the investment are left out of account in using only the common-sense "cost-reducing" criterion.

Such an omission will not matter if the amount of generated traffic is small, as it will be if the demand for transport is very inelastic. The reduction in transport cost will merely benefit existing traffic; virtually no new traffic will be induced to use the road. Such a rationalization for the "cost-reducing" criterion is, however, of limited use for developing countries, for it is clear that with many and perhaps most road improvements, one of the main objectives is to develop new land, new industries, and to open up larger areas to market influence. New development is important. We conclude, therefore, that the supposition of an inelastic demand is not a valid excuse for using only the "cost-reducing" criterion for investment.

Because it is uneconomic and even impossible to undertake road investment a little bit at a time, one must use the technique of measuring the "consumer surplus." Essentially this means calculating the amount of money people would be willing to pay for the new facility. Part of this value—and often the largest part—is obviously the cost reduction of existing traffic. But there is also the amount of money new traffic would pay for the facility. Since it is a hypothetical figure, this is difficult to measure. Conjectures about traffic flows generated by a reduction in transport cost are not an altogether satisfactory basis on which to erect investment programs.
But what are the alternatives? Clearly one might "allocate" the investment costs of the new facility to the traffic that uses the road. So the road authority would not supply a new road, or improve an old one, unless its costs would be covered by the additional revenue collected. But such a policy of charging the users involves large wastes, as we saw above. And it is not at all clear that such an investment criterion is less hypothetical, ex ante, than the criterion of consumer surplus. It involves similar judgments about the reaction of users to different levels of user charges and about the extent of generated traffic (if any!). Ex post there would be a "sufficient" condition that the investment was worthwhile (if the increase in revenue covered the cost). But by then the road is built, the resources are sunk, and the investment is a bygone; such historical knowledge is interesting and helps improve forecasting—but all ex post studies do that. Clearly the inefficiencies introduced by prices based on "investment costs" are not counterbalanced by any marked sharpening of investment criteria; on the contrary, the errors of pricing and investment would be compounding rather than compensating.

**Locating Plants and Farms**

It is sometimes alleged that, if road users are not required to pay the fixed costs associated with the provision of the highway, industry and agriculture will make inefficient decisions about the location of plant and farm. If the government bears a substantial fraction of the fixed costs of location, it may be thought that the plant or farm will be located so that it requires too much of the (allegedly) subsidized road transport—and (perhaps?) the government will be induced to overbuild the highway system.

All these conclusions are, however, wrong. If the road authority levies the economic user charges for the use of the road and if it employs the consumer surplus criterion for investment, there will be no inefficient location decisions and the authority will not "overbuild" the highways. And this will be true whether or not the government collects the "fixed" charges by a levy on the rents generated by the building or improvement of the road.

The authority will build the highway if the amount users would be willing to pay for the road exceeds the cost. But the sum of money they would be willing to pay is simply the increment in their profits (or rents). Consequently, the road authority will build the road if the profits (and rents) so generated exceed the cost.

Clearly, it does not make any difference to the location decision if the authority taxes away almost all of the additional profits; nor does it affect the location of business if the additional profits are left in the hands of the firms that enjoy the benefits of the road. The distribution of the additional profits (or rents) between government and business will in no way affect the location of industry.

**"Covering the Costs"**

One of the main differences between the system of user charges advocated here and those which are currently accepted by many transport experts is that there is no guarantee that with the proposed system the total amount collected in user charges would cover the total costs of the road network. With the currently accepted principles of user charges, the costs of the roads would be covered; the roads should not incur a deficit. And presumably, although this is rarely explicitly stated, road taxes should not be so high that the roads earn a surplus above costs. This is a version of the old public finance adage: "Each tub must stand on its own bottom."

The immediate appeal of a self-financing system of highways is obvious: it is consistent with superficial ideas of equity, and it is undoubtedly true that the public at large accept and even applaud the principle of the tub. When it gets down to the definitions to be used in practice, however, one finds that there is no principle by which one can define "costs of the road system," while even "user charge" is not easy to specify. This is not surprising because there is no logically consistent theory on which the "cover the costs" principle is based. It is a loose precept of social justice; it is not the consequence of a tight chain of consistent reasoning.

Nor is there any likelihood that the proposed system of economic user charges would result in a balanced road budget—however that may be "reasonably" defined. Clearly if a large fraction of the total vehicle mileage is incurred in congested cities then there is a good chance that the amount of revenue collected in user taxes will exceed the total costs of the highway system. If, on the other hand, a substantial proportion of the vehicle mileage is on uncongested highways it is likely that the road account will be in the red.

**Economic User Charges and Budget Deficits**

There is no overwhelming presumption that the economic user charges would generally give rise to large deficits on the road budget. But the only way to determine this is to calculate and see. Some rough and ready calculations for Thailand suggest that the introduction of very modest congestion charges would result in no substantial "deficit" on the road account. It must be emphasized, however, that such figures are exceedingly arbitrary and there are as many (and I suspect more) rules for making up a road budget as
there are transport economists. One should therefore not attribute any generality to such calculations.

Suppose, however, for the sake of the argument, that there were a deficit in the road budget as a consequence of adopting the proposed system of user charges. Would there be a good case for somehow increasing motor taxes so that the road budget was balanced? Is the tub really best standing on its own bottom?

No such simple conclusion can be drawn. The approach we suggest should be less doctrinaire and more pragmatic. Supposing that the system of user charges which we propose here does give rise to a deficit, then we ask—which would be the best taxes to raise in order to finance it? There is no presumption, from economic arguments, that the best taxes are those on motor vehicles or on vehicle-miles. All possible sources of tax revenue should be examined—and a decision should be based on a balanced rational evaluation of the consequences. Account must be taken of efficiency, the cost of administration, the effects on the distribution of income, and the various other considerations that bear on tax decisions.

Nevertheless, although there are no compelling economic reasons for balancing the road budget by levying road user taxes, there are good political and institutional reasons for introducing “earmarking” arrangements for the finance of roads. The main purpose of “earmarking” and “road-fund” arrangements is to insulate the decisions about road investment and motor taxation from the pressure groups and bargaining processes of politics. It may be hoped that some earmarking or road-fund approach will take most of the decisions about road investment out of the political arena, so that road investment is less likely to be used as a pawn or payoff in a political process. Decisions are more likely to be made rationally by an independent “public corporation,” such as a National Highways Board, which has an assured and earmarked income that cannot be used for any purpose other than that of investment and maintenance.

**Competition, Monopoly, and User Charges**

The proposed policy of levying economic user charges for the use of the road is, of course, derived from the presumption that other businesses in the economy are selling their goods in a competitive market. Indeed the simplicity of the theory and the sharpness of the results depend on the assumption of competitive conditions. But the competitive conditions which we have assumed could not conceivably exist in their pure form; in practice there are always elements of monopoly, “imperfect competition,” and strategic behavior. This does not mean that the “competitive theory” cannot be used in fact because its assumptions are violated; the assumptions of a theory are always “wrong.” What we must ask is whether the competitive theory predicts the behavior of the economy better than any alternative theory.

It is, however, natural to inquire into alternative models where the assumption of competitive conditions is dropped. What would be the best pricing policy for the road authority in a noncompetitive world? It must be immediately emphasized that there is no general answer to this question. One can only proceed pragmatically by asking what would be the major industries affected if the road authority varied its price (and its volume of investment). On the demand side the main effect would be felt by the complementary industries—truck and road passenger transport—and by competitive industries—such as rail and water transport. On the supply side, it seems that the main beneficiaries of a reduction in road investment would be those other industries that absorb the scarce factors used in road construction and maintenance; and it seems likely that these would not be concentrated in any particular industry but would be spread widely throughout the rest of the economy.

Obviously the industry that causes most (political) heartache is the railway. Usually railway charges are regulated by government and they may (and usually do) have only the most tenuous connection with marginal cost (suitably defined). Then the effect of road pricing on the railroads depends on (a) the amount of competitive traffic and (b) the relationship of rail rates to marginal cost for the traffic that would transfer from road to rail (or, conceivably, from rail to road). If there is little competitive traffic compared with the large amount that goes by road, then it is probably best to ignore it. If competitive traffic cannot be ignored, however, and if there is little substitution between transport and other goods, then one must determine the relationship between rail rate and marginal cost. There is no reason to suppose that the rail rate for this competitive traffic will be above marginal cost; in fact for much short-distance and medium-short-distance traffic regulated rates are often found to be much below cost. This would then suggest that the road authority should levy rates below the economic user charges.

In those circumstances where the regulating authority requires that the railway charge prices above marginal cost for traffic that is competitive with road, there may be a case for raising road charges above the marginal costs. The reader may well wonder, however, whether it would not be wise to concentrating on reforming rail charges rather than distorting road user charges to fit into the rail world; a better approach might be to make the railway pricing policy more efficient.

One might conjecture that, if for some reason the railways are required to “cover their costs,” the road users should be required to cover theirs. No such
conclusion can be drawn from the economic analysis of road-rail competition. It may, or it may not, be a good idea to make a profit or loss on the road system as a whole. The issue is influenced only slightly by the railways being required to break even. What really matters are the effects of budget constraints on prices when there are competitive services available; and it will be in the interests of rational rail management to distort prices little from cost when there is a close competitor.

No clear-cut conclusions emerge from this analysis. But theory does give us some guidance; it does provide a filing system for approaching the real world. We can estimate the competitive rail traffic and one may make judgments about the relationship between marginal cost and price. But there is no simple rule to be followed: the desirable road price might be above or below the marginal cost.

The Guidance Offered by Theory

The general conclusion of this study is that the best simple policy is to begin with the economic user charges, or the short-run marginal cost. This should be the basis of all road charges. Then judgments about departures from the economic user charges must be made in the light of (a) the administrative and other costs involved; (b) the need for revenue and the alternative sources; and (c) the competitive structure of the economy. Each departure should be justified by arguments about the consequences. But the bedrock of road user charges remains the economic user charge.

The Unit of Cost Analysis

Some exponents of the theory of road pricing have argued that the long-run marginal cost is the appropriate value for fixing prices. In this study, however, we have stressed that the short-run marginal cost and not the long-run marginal cost is always the appropriate concept for pricing purposes. It is easy to show that the long-run marginal cost is efficient only when three conditions are met. The first requirement is that traffic be growing and expected to grow, so that there is no question of simply making the best use of a road that is too big; all roads must be enlarged. The second condition is that there be no indivisibilities and joint-product relationships in the supply of roads. The third condition is that the government always carry out the efficient investment program, never lag behind in building new roads when they are desirable, and never make mistakes by overbuilding.

If these three conditions are present, the long-run marginal cost will provide an appropriate measure of the efficient user charge. It is important to see why under these conditions the long-run marginal cost pricing policy is correct. For it is only when these three requirements are met that the long-run and short-run marginal costs are equal. In other circumstances they will generally be different. Thus the long-run marginal cost is useful as a criterion for user charges only when it is equal to short-run marginal cost. When the long-run differs from the short-run cost, the long-run concept is the wrong basis for pricing policy. The short-run marginal cost is always the appropriate value at which to fix the user charge.

Any concept of cost must be associated with the decision to which it is related. The concept of cost used throughout the whole of this study comprehends solely the escapable cost as one course of action is selected rather than another. In any well specified situation there are a number of feasible alternative courses of action; and also a set of circumstances which one cannot change. Taking the latter as given, we calculate the value of the resources used up if we adopt different courses of action. There is therefore an infinity of concepts of cost—but each one is uniquely related to the decision specified in each situation.

Which concept of cost is relevant for the user charge policy of the road authority? For this purpose we must inquire into the role of price in the budgets of consumers and firms that use the highways. Clearly each user takes the price (user charge) as given and “buys” highway services so that the additional satisfaction derived from the last unit of highway service just equals the price he has to pay. Since the consumer of highway services can buy units of highway services as small as he likes, the relevant unit for pricing purposes is the “small” increment in road service which the consumer may buy. The appropriate cost concept is therefore uniquely defined as the marginal cost of supplying an additional (small) unit of the road service—and the marginal cost is “the value of the resources which would be saved by not supplying a (small) unit of road service.”

It is quite wrong to imagine that this concept is arbitrary. It is not. It is determined by the nature of the “commodity”—it is possible to use a road in very small units at a time—and by the fact that the de-

9. It is easy to sketch (but perhaps hard to imagine) the conditions under which the relevant pricing unit would be much larger. Suppose, for example, that the road authority decided to sell road journeys in “blocks” of 1,000 vehicle-miles, and that subcontracting (or retailing road services to third parties or trading in any units smaller than 1,000 vehicle-miles) was declared to be illegal with very high (prohibitive) penalties. Then the only possible unit of contract is the 1,000 vehicle-miles of road service, and this is the relevant unit for the analysis of cost. But it is misleading to talk about the marginal cost of these large units. We shall analyze all cases with “sticky” prices in the next subsection.
decisions of consumers in allocating expenditure between road services and other goods take this divisibility into account. The road user decides on the little more or the little less, and so it is the cost of the little more or the little less that is relevant for determining the price.

One of the greatest sources of confusion in the analysis of road costs is the confounding of discontinuities on the supply side with those divisible and marginal decisions on the consumption side. The decisions of the road authority encompass both the supply of the road and of road services. What about the additional cost of supplying the road? Clearly, if we lumped all the road users together and treated the collective as a contracting party there would be a correct criterion for investment by asking whether it would pay a sum of money greater than or equal to the cost of the road. (This is, of course, simply a version of the consumer surplus criterion, but we leave this aside for the moment.) This is no basis for determining the price of road services however. The collective does not buy road services, individuals do. Thus, as we argued above, the consequence of the decision of the consumer, in buying a little more or a little less, is the relevant concept for the calculation of cost, and thus the price of road services. The road authority makes investment decisions in accordance with the evaluation of the collective; but it sells road services to individuals, not to the collective. Again we conclude that the marginal cost of a small unit is the relevant concept for pricing.

The Definition of the Short Run

Characteristically we think of price as the value that moves to “clear the market.” Price adjusts to the level where supply equals demand. Movements of price, we assume, do not involve any real costs; there is no cost involved in “relabelling” the price tag. This picture of frictionless price adjustments corresponds closely to reality in markets for certain kinds of road haulage, taxis, and buses in some countries of the developing world. Under competitive conditions no seat would be vacant and no capacity unoccupied unless the price dropped to zero; then it would be a free good. Price will move costlessly and swiftly to the value that clears the market.

In the instantaneous run (or the very very short run) the price of road services is determined by the demand in conjunction with the existing capacity. Since capacity at any instant of time is given, and since bargains can be concluded costlessly, no resources can be saved by supplying one less unit of service (for example, by not filling a seat on the bus), nor is it possible to supply any more (no standing passengers permitted by law). The maximum volume of services is given in the instant—and price must perform the task of allocating the services among the users who are prepared to pay.

This model is the same as the simplified model of a road which we developed earlier; the only difference is that the instantaneous operating cost may be taken to be almost zero. The solution there was to charge a price of marginal cost including the rent necessary to equate demand and supply. The marginal cost used for pricing purposes was carefully defined to include the rent. In the instantaneous run, it seems that the rent will comprise virtually the whole of the marginal cost, as we have defined it. Price, duly fixed at every instant at marginal cost, will move immediately to clear the market if there is excess demand or excess supply. The appropriate concept of cost for pricing purposes is, therefore, the instantaneous marginal cost (including rent), or, to use the usual terminology, the very short-run marginal cost.

With this model of a frictionless price, we might now consider a longer period where some obligations are escapable; the vehicle owner can decide whether or not to take the vehicle journey. Of course price is still determined by the instantaneous market conditions. The vehicle owner, however, now takes into account the alternative use of the resources which he may commit to road transport for the period. If it seems profitable to put more of his resources into road services, the vehicle owner will run more vehicles, and the increased service capacity will depress the price.

Of course this frictionless price model is of limited use for practical application in road pricing. Frictions do exist. But it is important to see that the “pure” frictionless case does require that price be fixed each instant at the very short-run (instantaneous) marginal cost (including rent). The instantaneous “period” is defined automatically by the frictionless character of price; there is then no problem of defining appropriate periods for pricing purposes. It is also worth stressing that the completely flexible frictionless price, moving effortlessly and costlessly to the value that “clears the market,” is the peak of efficient pricing. It represents a standard of aspiration which is useful to keep in mind when developing more mundane and workable schemes. It is, indeed, important to get as much “built-in” flexibility as one can into administrative schemes of prices.

The main frictions in road pricing arise from the political and institutional processes that determine user charges. To discuss their impact would require a survey of the political institutions of each country. Instead we shall briefly discuss some of the principles.

10. I am assuming here that the cost of filling a seat is zero.

11. What we have elsewhere called variable maintenance costs are committed and fixed in the instantaneous run.
of pricing under conditions of friction. Obviously the most straightforward approach is to include a "cost-of-changing-the-price" term in the model. The formal conclusions of the model are easy to predict; it would sometimes pay to waste some road services rather than change the price, and, up to certain levels, it would pay to tolerate queueing and congestion rather than increase the user charge. It would be worthwhile to vary the price only if a large change were to be made. Small changes would be too costly and would be ruled out. Prices (or user charges) must then be moved only in steps; the property of infinitesimal adjustments disappears. Price is "sticky."

With this model it is necessary to deal with "blocks" of services. If the price is reduced then, since it must be reduced by a "large" amount, the expansion of services must also be more than infinitesimal. The cost of changing the price must be introduced into the cost calculations for the purpose of fixing a new price. The discontinuity introduced by the "cost-of-changing-the-price" term is analogous to the "all-or-nothing" investment decision. This implies that we must analyze the cost of a lump of output, that is a block of road services. In principle, this also involves the calculation of consumer surpluses—just like the lumpy investment decision. The gain from reducing the price, when there has been a fall in demand, will be reckoned in terms of the amount which users would be willing to pay for the additional services; against this gain is to be set the cost of changing the price and the cost of supplying the additional services.

A diagrammatic illustration (figure 4-1) might help clarify this point. Let us begin with the demand curve $D_o$, the marginal cost curve $MC$, and price $P_o$. Suppose that demand falls to $D_1$. If the price were frictionless it would fall to $P_1$. But suppose there is a "cost-of-price-change" involved; then there is a choice between holding the price at $P_o$ and reducing output to $q_1$, or lowering the price to $P_1$ and expanding output to $q_2$. Consumers therefore are better off if the road price is reduced by an amount represented by the triangle ABD. The producer surplus is given by the triangle BCD. But against these two gains from the price move we must set the cost of changing the price. If the cost of reducing the price is sufficiently large it may offset the gains to be obtained.

We might at this stage ask whether the "costs-of-price-changes" can be reckoned to be important enough to dominate the analysis of user charges. No definitive answer can be given until there has been an adequate study of the facts, but some points seem clear. There is a great difference between a change in the structure of prices, including changes in the relative prices of different types of roads, etc., and changes that involve simply predetermined variations in user charges. The former cost a lot, the latter are almost free. Really major changes in user charge policy are almost once-and-for-all changes and one would not expect them to be repeated frequently. But it does not seem difficult or expensive to introduce some "built-in" flexibility into user charges. They might well be varied at trivial cost under rules of procedure; for example, the price on an urban tollway could be raised for the rush hours and reduced for off-peak periods. Thus it seems likely that, for normal price changes, the cost of variation will not be so high that it dominates the analysis. But it is important to check these conjectures against the facts.

4.8 Port Pricing

One of the main themes of this study is that an appropriate basis for pricing port services is the marginal cost of supplying the services. One of the first

---

12. Formally this could be done by using a dummy vehicle $z$, which takes a value 0 if there is no change and 1 if changed; so the cost of the price movement (if any) would be represented by $zx$ where $x$ is the cost if the movement is actually made.

13. Some exponents of cost analysis talk about the marginal cost of a block of road services. This is, however, misleading since marginal analysis applies only to infinitesimals and the solution of "all-or-nothing" pricing proposals involves much more complex analysis and calculation.

14. The consumer gain of $P_0 P_1 BA$ is, of course, offset by the same producer loss.

tasks is to develop the principles of marginal cost pricing in their applications to ports. The background to marginal cost pricing has been described in some detail in its application to highways, and this survey is restricted to the highlights.

Short-run Marginal Cost Pricing and the Consequences

The argument for marginal cost pricing is that, under conditions where the rest of the national economy behaves (or is planned to behave) as if perfect competition existed, the price should be set equal to the cost of the resources absorbed in producing the service. If the price exceeds the cost, too little of the service will be produced, and resources will be used inefficiently to produce other commodities and services. Similarly, if the price is below the cost, too much of the service will be produced relative to the quantity of other goods.

But what marginal cost is appropriate to be considered here—short, long, medium, or some combination? Only a conditional answer can be given to this. If it is assumed that it costs very little to change the price, then the short-run marginal cost is an appropriate basis for pricing. If, however, the price is costly to change (or administratively difficult, which amounts to the same thing), then a somewhat longer-run marginal cost is the proper criterion. The constraints on the movement of prices (or the shadow price of such constraints) should be the only reason to switch from the short-run basis of marginal cost to the longer-run basis.

Whether short-run marginal cost is less than or exceeds the long-run marginal cost depends, among other things, on the existence of returns to scale in the port industry and on whether the capital stock and equipment is adjusted to the level of output to allow production at least cost. First, suppose that the demand conditions for future traffic levels are known exactly and that the port authorities have invested wisely to ensure the best technology for their scale of operations. This means that the port authorities are conducting their business at minimum cost and that the average long-run cost is the same as the average short-run cost.

But short-run and long-run marginal cost (SRMC and LRMC) will be equal only if there are constant returns to scale for the levels of traffic for which the facilities have been built, as can be seen in figure 4-2. Pricing the services at short-run marginal cost will attract just enough traffic to cover the port's cost. For traffic levels substantially lower, however, the port facilities would be underused and the short-run marginal cost would fall below the long-run marginal cost. For these low levels of tonnage there would be economies of scale. The general result, therefore, is that with no economies of scale, if the port has the optimal level of traffic, the long- and short-run marginal costs are equal. Furthermore, if there are economies of scale, levying a price of $SRMC \neq LRMC$ will give rise to deficits.

Now assume that the port authorities are not perfect administrators and that, being perhaps human, they make mistakes. Thus the wharves and equipment are not always appropriately adjusted to the level of traffic, perhaps because tonnage grew at an unanticipated high (or low) rate. Suppose, for example, that the demand turned out much larger than the authorities anticipated. Thus the short-run marginal cost will be high, reflecting the scarcity of port capacity, and well above the long-run marginal cost. If the port tariff were fixed near the value of long-run marginal cost, inefficient congestion would result. The tariff should reflect the scarcity value of the port's facilities as measured in the short-run marginal cost; then congestion would be virtually eliminated.

Suppose, on the other hand, that the authorities overbuild their ports. Then the short-run marginal cost is well below the long-run value. But if long-run cost were levied, the use of the underutilized port would be discouraged; thus the short-run is the appropriate value for the tariff.

The overall result, therefore, is that the long-run marginal cost is a suitable basis for pricing only when it happens to be the same as the short-run marginal cost, and this will occur only when the authorities predict their output and capacity correctly. However, the short-run marginal cost is always an appropriate base for pricing, irrespective of the efficacy of the
Marginal Cost Pricing Principles in a Noncompetitive Environment

The arguments for marginal cost pricing spring from the assumption that the price of a resource measures its marginal cost. This condition may arise in a capitalist economy if there is perfect competition in other activities. Marginal cost pricing may be obtained in a socialistic or dirigiste economy by fiat of the authorities who regulate prices. But both are ideal states—like perfect vacuums or perfectly accurate measures; in fact neither in a socialistic nor in a market economy can marginal cost pricing be precise. Mistakes will be made, monopolies will arise, taxation will distort prices, political pressure will be brought to bear, and so on. The question is whether a rule of pricing that is strictly relevant only when ideal marginal-cost-equals-price conditions are present can be used also in the mundane world of mistakes and imperfections.

Figure 4-3. Cost and Output with Discontinuities

Note: Definitions as in figure 4-2.

To answer this question it is necessary to consider some alternative principle. As a substitute for marginal cost pricing, the only really serious candidate is average cost pricing. (For this discussion the combinations of the two which are characteristic of two-part tariffs are ignored.) It is true that when marginal cost and price differ in the rest of the economy, there is no simple rule for adjusting the marginal price of the port. In particular, even though prices are above marginal cost in the rest of the economy, it is not known whether the port tariff should be above or below marginal cost. Nevertheless, if the competitors for the specific resources used by the port (for example, capital) charge prices that are considerably above marginal cost, then it may be wise to raise tariffs above marginal cost in the port; otherwise the port may expand too much and take too many of these scarce resources from the other industry. (However, this supposes that the resources are not producible, which is not true in the case of capital.)

As far as competition for resources is concerned, decisions can only be made on the evidence presented by the particular circumstances. Much depends on whether the competing demands for the resources used by the port come from commodities or services that are competitive with or complementary to port services. In particular, if the competing buyer of capital produces complements to the port's output (for example, road transport to or from the port), the most advantageous pricing rule would be for the port to set its prices below its marginal cost and vice versa for the case where the competitor for the port's resources is the monopolistic producer of goods or services that are substitutes for the port's output (for example, another national port or import-
competing production in general). As a practical matter it seems that the main competitor for resources is probably road transport, including the provision of highways. The output of road services usually complements that of port services—but they are often produced in conditions of energetic competition. Other competitors for resources may be domestic construction industries. Here again competitive conditions are usually observed. Therefore, it is deduced that there are no substantial arguments for pricing below marginal cost.

Whether prices should be levied substantially above marginal cost depends, among other things, on the structure of competition among the customers of the port. The most interesting case here is the liner conference and its relation with the traders who, in turn, comprise its customers. These considerations involve many problems about the reaction of the shipping conferences to changes in their cost and will have to be pursued at length elsewhere.

The alternative principle of average cost pricing is difficult to define in a unique way when there are many services performed by broadly the same activities. The common cost of central activities must be allocated arbitrarily among the different services produced by the port. There is no unique measure of average cost for the multiproduct firm. In many costing studies common cost is allocated according to weight, or some other measure of volume, or perhaps according to some measure of what the traffic will bear, or some combination of the two. But of course the allocations are not of interest for pricing purposes except as approximations of the avoidable cost of providing the services.

**Pricing in Ignorance of Cost**

The final question in this discussion is the issue of imperfect knowledge and marginal cost pricing. One of the recurrent themes of many port studies is that the authorities do not know what their cost is. And if they lack knowledge of their accounting cost, however defined, it seems at least likely that they do not know their marginal cost. This is particularly important since marginal cost is a useful tool for efficient management.

Some knowledge of cost is necessary for a port authority to discharge its most basic obligations. These include the efficient conduct of the business such as buying properly, preventing and rooting out fraud, and adopting the best techniques and administrative practices. Knowledge of marginal cost, both in the short and long run, is very useful for making the appropriate decisions on investment to meet new or growing demands for port services.

In practice many of the components of marginal cost can be adduced from the financial accounts of the port. But the financial accounts of some port authorities contain an assortment of transactions which are difficult to disentangle, and in some cases the port accounts are inextricably mixed up with other agencies. To determine marginal cost it is then necessary to instigate specific surveys of port activities. Only then can the alternative investment programs be judged on their ability to meet future demands.

But the appropriate pricing for port services need not wait for such time-consuming and costly studies; the tariffs can be improved immediately. In practice the first step is to levy charges so that the use of the port is either increased or reduced to what is regarded in port economics as the efficient level. The efficient level of port use can be derived by observing ports, perhaps in other countries, where the management is efficient and where the port is operating at the "best" level. This comparative standard will provide a good enough first approximation for pricing purposes. Then the port authorities need to estimate the reaction of the demand for port services to changes in the price. Essentially, instead of calculating marginal costs, the problem is reduced to estimating the elasticity of demand for port services.

This argument applies to the general level of port prices. In addition, the structure of port tariffs must be correct so that the price of each service reflects the cost involved in supplying it. This principle should be applied to the separate services and activities of the port. Thus, the price of storage should be fixed so that the facility is used as near to its physical capacity as is efficient, and similarly quay charges should be varied until the utilization of the quay is appropriate. (This may involve very high congestion charges on some facilities.)

Strictly speaking, the method outlined above is only appropriate for pricing facilities whose use does not incur avoidable cost. In that case the rule of pricing for the full use of capacity is appropriate simply because the definition of full capacity is likely to be that volume of output at which marginal cost starts to rise rapidly. By extension, the rule may apply also to pricing facilities which incur little variable cost in their use. The rule is not appropriate, however, for services which have a substantial cost that varies with the amount of the service produced.

But port authorities are never totally ignorant of cost. Usually, they have only hazy or erroneous notions about fixed cost, but they generally can formulate much more accurately the magnitude of variable (or avoidable) cost. If the variable cost of a service can be guessed with some confidence, marginal cost can be derived from it; in that sense, marginal-cost pricing demands less information about cost than does pricing according to average cost. Indeed, since marginal cost can be approximated as the change in variable cost as output changes, a clear
The idea about absolute variable cost is not required. To ascertain changes in variable cost may require study, but less than a study of the full cost for the port.

Last, and perhaps most realistically, the port authorities may be able to quantify only part of the variable cost of a service. Such a port will then fix its prices substantially above that known element of variable cost to recoup unknown cost elements. If the authority must cover its cost, it will seek a total revenue equal to or above total variable cost plus total assumed amortization of assets. So far as the cost of rendering a particular service is concerned, the port is most likely to know the direct variable cost and to feel ignorant about indirect items. These are frequently the costs of services or supplies used in common by several activities. In that realistic case, again, part of the marginal cost of rendering the service can be estimated from what is known. The marginal-cost pricing rule then yields the important principle that the price charged for the service must not be less than this ascertainable marginal cost.

Concepts of Marginal Cost

Some of the preceding readings raised questions about the interpretation and measurement of marginal cost. The next reading explores this topic in more detail, with special reference to public utility pricing.

4.9 Alternative Definitions of Marginal Cost

In the following sections marginal, or incremental, cost is defined in four slightly different ways. The definitions are similar in that they are forward looking, that is, they consider only future costs and future output. The definitions differ in the extent to which they stress the importance of short-run as opposed to long-run costs, operation as opposed to investment costs, and changes in consumption in different time periods. In consequence, when used for pricing purposes they vary in the extent to which they focus on short- versus long-run allocative efficiency, and the extent to which they attempt to minimize price fluctuations in the presence of lumpy investments. The four definitions examined are:

1. "Textbook" marginal cost (TMC)
2. "Textbook" long-run incremental cost (TLRIC)
3. Present worth of incremental system cost (PWISC)
4. Average incremental cost (AIC).

"Textbook" Marginal Cost

Marginal cost can be defined as the first derivative of total cost with respect to output, where TC = f (output) is a continuous single-valued monotonic function. When, however, a total cost function is discontinuous (investment is lumpy or cost records are kept on the basis of yearly changes and not by changes for each additional unit), the above definition is not strictly applicable. In such cases there is considerable ambiguity in economics literature about how to choose the appropriate magnitude of output change, and about the appropriate cost which should be attributed to that output. It is possible, for example, to argue that the total cost of an additional lump of investment should be attributed to the very last unit of output. Such an interpretation, however, can be opposed on the grounds that additional investment is justified when consumers show their willingness, over a given period, to assume the financial (and presumably economic) burden of additional capacity investment measured over the same period.

This burden can be defined as the payment that must be made to cover the operating and amortization costs of the additional capacity. In practice it is inconvenient to use instantaneous increments and amortization schedules arranged in minute time periods, and in our definition of TMC we use periods of one year for these purposes.

Of course, if cost-benefit analysis is feasible the problem of deriving an investment-signal price can be avoided. Then, in cases when investment is lumpy,
price is set equal to short-run marginal cost, and when excess capacity nears exhaustion a cost-benefit exercise is undertaken to see if additional investment is justified. If it is justified, the investment is undertaken and price is then set equal to the new short-run marginal cost. If, however, investment is not shown to be justified at that time, presumably price would be allowed to rise to allocate the perfectly inelastic supply. In this circumstance the behavior of price would correspond closely to the TMC definition outlined below. Of course, the difficulties of benefit measurement related to public utility output generally preclude the option of short-run marginal cost pricing with cost-benefit analysis justifying additional investment.

The TMC definition of marginal cost examined in this paper generally reflects micro-economic pricing theory, with the modification that each increment is taken to be the change in output that occurs during one year. As a result, TMC makes use of two concepts: short-run marginal cost (SRMC), which reflects increments in operating and maintenance costs brought about by increases in output, and marginal capacity cost (MCC), which reflects increments in capital expenditures (capacity) that are necessary to increase output. Since these cost and output increments are considered only one year at a time, the TMC definition reflects a relatively short time horizon.

TMC is defined as:

$$TMC_t = SRMC_t + MCC_t,$$

$$= \frac{R_{t+1} - R_t}{Q_{t+1} - Q_t} + \frac{r_l}{Q_{t+1} - Q_t},$$

where $t$ is Year for which TMC is being calculated,

$R_t$ = Operation and maintenance expenditures in year $t$,

$Q_t$ = Water produced in year $t$,

$I_t$ = Capital expenditures in year $t$,

$r$ = The capital recovery factor, or the annual payment that will repay a $1 loan over the useful life of the investment with compound interest (equal to the opportunity cost of capital) on the unpaid balance.

Since there could be a vertical gap between SRMC and TMC when capacity is reached, for purposes of our subsequent analysis the increment in consumption used to calculate the MCC part of TMC is that for the year immediately following the expenditure on new investment and not for the year before the expenditure.

With a lumpy investment stream, TMC for the years in which capacity expenditures take place reflects both SRMC and MCC; during years in which no capital expenditures take place TMC equals SRMC only. Charging according to TMC, therefore, involves significant fluctuations in price, the signal to invest in additional capacity being given when output is at capacity and the price paid by water consumers, for instance, is equal to short-run marginal cost plus annual equivalent marginal capacity cost (MCC).

"Textbook" Long-run Incremental Cost

Use of the TLRIC definition of marginal cost emphasizes the need to give investment signals to present and potential water consumers at the expense of some loss in short-run allocative efficiency. As normally presented in economics literature, the forward-looking time horizon of TLRIC does not, however, extend beyond the next investment. TLRIC is defined as follows:

$$TLRIC_t = \frac{R_{t+1} - R_t}{Q_{t+1} - Q_t} + \frac{r_l}{Q_{t+1} - Q_t},$$

where the notation is identical to that used for TMC.

$$r = \frac{l i (1 + i)^n}{(1 + i)^n - 1},$$

where $l$ is the investment cost, $i$ is the appropriate interest rate, and $n$ is the useful length of life of the investment (assumed to be thirty years).

21. A problem with the use of this definition in subsequent analysis is that if the price elasticity of demand is different from zero (and there is considerable evidence to suggest that in most cases it is), the consumption (and output) of water would increase more rapidly (assuming demand at a given price is growing at some compounded rate) during years when there is large excess capacity, and less rapidly in years when the utility is approaching capacity and short-run marginal costs are rising. As a result, the TMC estimate of SRMC, which we later assume to be increasing at the same rate as output, could in reality be an underestimate as output nears the capacity of the utility.

22. Of course TLRIC could be redefined to look at the average of the next several investments or at one particular "representative" future investment.

18. This alternative might be appropriate in the evaluation of certain highway projects. See A. A. Walters, The Economics of Road User Charges (Baltimore, Md.: Johns Hopkins University Press, 1968).

19. Our definition of TMC is therefore somewhat arbitrary with regard both to the magnitude of the increment in output and the application of an amortization factor to capital costs. Even the textbook definition, as we define it, therefore makes use of some "smoothing." Definitions that fail entirely to do so, we would argue, do not deserve to be considered seriously as practicable alternatives.
with the exception that \( k \) denotes the year in which the very next major investment expenditure is completed. As a result, during the years \( t \) through \( k \) the term \( \frac{I_t}{Q_{t+1}} - \frac{I_k}{Q_k} \) remains constant, reflecting the annual equivalent of marginal capacity cost for the next lump of investment. In year \( k + 1 \), after investment has taken place in year \( k \), \( k \) is redesignated to be the next year in which a large investment is completed, and, again, over the period \( t \) through \( k \) the term \( \frac{I_t}{Q_{t+1}} - \frac{I_k}{Q_k} \) remains constant, reflecting the annual equivalent of marginal capacity cost for the next lump of investment. For years in which investment takes place, TLRIC will equal TMC.

The TLRIC definition jumps from investment peak to investment peak with price changing immediately following an investment to reflect the incremental costs of the next big capacity investment that will have to be incurred. This method, or close variants of it, has been most frequently used in the electric power sectors where the need for forward-looking investment signals is critical, and where because of the existence of interconnected nationwide grids, lumpiness in investment is less of a problem.

**Present Worth of Incremental System Cost (PWISC)**

PWISC is another method of defining marginal cost that emphasizes the necessity of providing consumption and investment signals that reflect the magnitude of forthcoming investment. As presented by Turvey, the PWISC definition does not, with the exception of replacement costs, look beyond the next lump of investment, and therefore also tends to ignore the effect of increasing or declining unit costs associated with subsequent increases in output and consumption. For purposes of water supply pricing PWISC is defined as: the present worth of the increment of system costs resulting from a permanent increment in consumption at the beginning of year \( t \) minus the present worth of the increment of system costs resulting from the same permanent increment in consumption starting at the beginning of year \( t + 1 \). Algebraically, this can be expressed as:

\[
P_{t, t+1} = \frac{R_{t+1} - R_t}{Q_{t+1} - Q_t} + \frac{I_{k+1}}{(1+i)^{k+1} - (1+i)^{t+1}}
\]

\[
+ \frac{I_{k+29}}{(1+i)^{k+29} - (1+i)^{t+1}} \left( \frac{1}{(1+i)^{30} - (1+i)^{t+1}} \right)
\]

where the notation is similar to that used for TLRIC in that \( k \) again denotes the year in which the very next large investment expenditure takes place (the year in which the system reaches capacity), where \( i \) is the opportunity cost of capital, and where the useful life of investment is assumed to be thirty years.

**Average Incremental Cost (AIC)**

The AIC definition of marginal cost represents an attempt (1) to compromise between short-run allocative efficiency goals and the need to signal the justification of investment in additional capacity, and (2) to look beyond the traditional economic definition of the long run by including all future investment costs during a specific time period; usually ten to fifteen years would be the maximum period for which reliable data would be available. In looking beyond the next increment in capacity, AIC makes different assumptions about the proportion of the investment that must be paid at one point in time in order to reveal consumer willingness to pay, and about the relevant magnitude of the next increment in capacity, which is invariably difficult to specify, particularly in large and complex systems in which many investments (some of which produce joint products) are taking place simultaneously.

Basically AIC assumes that when investment is lumpy, marginal capacity cost (MCC) can be estimated as:

\[
MCC = \text{Present worth of the least-cost investment stream} / \text{Present worth of the stream of incremental output resulting from the investment.}
\]
as defined above, this is an indication that the benefits of the investment program, which can include one or more projects and which would be measured in cost-benefit analysis as the present value of total revenue \((P \times Q)\), are at least equal to the costs, measured as the present value of the stream of costs to be incurred through time. The average incremental cost estimate that we simulate is calculated by discounting all incremental costs that will be incurred in the future to provide the estimated additional amounts of water that will be demanded over a specified period, and dividing by the discounted value of incremental output over the period, that is,

\[
AIC_i = \frac{\sum_{t=1}^{T} \frac{(R_{t+1} - R_t) + I_{t+1}}{(1 + i)^{-t}}}{\sum_{t=1}^{T} \frac{(Q_{t+1} - Q_t)}{(1 + i)^{-t}}},
\]

where the notation is similar to that used previously except that \(T\) is the number of years for which water expenditures and attributable output are forecast and over which price is being smoothed. Theoretically, the interest rate \(i\) in the numerator of \(AIC\) should be set equal to the opportunity cost of capital, while that in the denominator should be equal to the time preference rate for consumption, reflecting the assumption that consumption today is more valuable than consumption in the future.

The \(AIC\) definition gives marginal cost estimates that smooth out lumps in expenditure streams while at the same time reflect the general level and trend of future costs that will have to be incurred as water consumption increases. Where unit costs are lumpy, the \(AIC\) estimate will never be equal to the \(TMC\) estimate; it will always be above \(SRMC\) and usually below \(TLRIC\), and in relation to \(TMC\) will therefore tend to discourage "justified" consumption when there is excess capacity and to provide premature investment justification signals as capacity reaches full utilization.

The Choice of Definitions

In the abstract, marginal cost (the change in total cost brought about by an infinitesimally small change in output) is a simple concept. In practice, however, where we are dealing with changes in both costs and output through time, many different forward-looking definitions of marginal cost are employed. The version actually selected may depend upon:

1. The size of the increment in output that is being examined;
2. the length of the time horizon which is considered to be relevant, or the period of time for which significant excess capacity will exist;
3. the desired emphasis on near-term allocative efficiency versus longer-term resource allocation;
4. the rapidity with which changes in technology or economies of scale are expected to affect significantly production costs;
5. the extent to which relative price stability through time is thought to be necessary; and
6. revenue implications.

The four definitions examined in this paper generally cover the spectrum of ways in which most of these factors can be traded off against each other. Of these methods, \(TMC\) adheres most strictly to the real marginal costs that are actually incurred at any one point in time, although as noted previously the actual magnitude of \(TMC\) is dependent on assumptions concerning the appropriate output increment and the proportion of capacity cost that must be covered for consumers to reveal their willingness to pay. Furthermore, even with the moderate "smoothing" that the \(TMC\) definition incorporates, in view of its implications for price fluctuations when investments are lumpy, its practical application has been questioned by many economists.

To remedy this price fluctuation deficiency, and to address the fact that present and potential water consumers have imperfect information about future water prices, it is necessary to incorporate marginal capacity costs in prices even when investment in capacity is not imminent. Several alternatives for doing this have been proposed. While recognizing the concept of marginal capacity costs, \(PWISC\) does so in such a way that considerable price fluctuation remains, price falling significantly following an investment and rising gradually to approximate \(TMC\) when the next investment is due. \(PWISC\), \(TLRIC\), and \(TMC\) therefore adequately signal the need for investment in capacity, but only \(TLRIC\) avoids the significant drops in price associated with the other two methods at times of temporary excess capacity. Indeed, \(TLRIC\) implies relatively stable prices between
investments, significant changes taking place only immediately after an investment. 28

AIC is distinguished from the other three definitions in that it takes a longer view of costs, looking beyond the next increment in capacity. This can be particularly important when rapid technological change is taking place or significant scale economies are being experienced. AIC also has the attribute of avoiding severe price fluctuations although it does not adhere closely to TMC either at capacity points (as do TLRIC and FWISC) or during periods of excess capacity. It is thus essentially a compromise solution, neither adequately (in a textbook sense) signalling the justification for any one specific investment in a program, nor corresponding to the textbook SRMC ideal in the short run.

Other definitions that are variations on the above basic themes are, of course, possible. We have specified and simulated a number of these, including several that were derivatives of AIC and FWISC, TMC, and AIC and TLRIC. Our conclusion is that for water supply pricing, the four basic definitions indicated above adequately represent the spectrum of alternatives.

4.10 Water Supply Pricing

Experience suggests that in the public utilities field in general, the only way in which the minimum economic worth of investments can be determined is by giving consumers themselves the chance to let authorities know how much they value the service concerned, by being charged a price that reflects the full economic cost of supply.

Some Basic Principles

We shall outline the role of economic pricing theory in the water supply field, and show that, while rigorous application of the theory may not always be feasible in rural areas, the principles remain highly relevant. Indeed, we contend that, combined with the financial and administrative arguments for recovering costs from consumers, recognition of the concept of consumers' willingness to pay as a guide to resource allocation is absolutely essential to the achievement of any noticeable improvement in the rural water supply situation in the developing world.

28. No change takes place in the constant cost case.


ECONOMIC EFFICIENCY AND MARGINAL COST PRICING

An important benchmark by which policies relating to water supply may be judged is the contribution those policies make toward economic efficiency. An efficient policy may be defined roughly as one which maximizes the net benefits accruing to a community from a given course of action, with no consideration paid to the way in which those benefits are distributed within the community. A proposition stemming from this definition is that the price of any service or commodity supplied by a public body should be equated to the cost of producing an additional unit of it, or in other words, to its marginal or incremental cost. If consumers are willing to pay a price that exceeds marginal cost, it means that they place a value on the marginal unit consumed at least as great as the cost to the rest of society of producing that unit; output and consumption should therefore be expanded when system capacity is reached. If, on the other hand, the market clearing price is less than marginal cost, it can be assumed that there is oversupply of the commodity; the cost of additional output exceeds the benefits.

Whether or not a policy is thought to contribute to a movement toward efficiency will, of course, depend upon the commodity whose benefits the analyst is interested in maximizing. Having determined the relevant target group of people, he must distinguish between purely accounting costs and real (or economic) costs incurred by that group. The former costs, which might include repayment of past loans, simply represent a transfer of income within the community. Efficiency in resource allocation dictates that these "sunk costs" be ignored for pricing purposes, for they represent no net loss, or avoidable cost, to society as a whole. On the other hand, the resources employed in the construction and operation of a particular project represent, at the time of employment, real costs in terms of opportunities forgone elsewhere. The price charged for the good or service concerned should clearly incorporate recovery of such costs if they are incurred as a result of additional consumption.

The principle outlined here is straightforward. There are a number of difficulties, however, regarding both its practicality and its desirability that surround its implementation. Other problems such as capital indivisibility, financial constraints, and the use of shadow values arise in all applications of marginal cost-pricing principles, while important, they are not peculiarly rural water supply problems and are only briefly discussed in this section. The specific problems of measuring water consumption and of supplying the rural poor are dealt with separately.
CAPITAL INDIVISIBILITY

The foregoing remarks suggest that a distinction should be made, for water supply pricing purposes, between those costs that are a function of consumption and those that are not. Ambiguity in the definition of marginal cost arises where capital indivisibility (or "lumpiness") is present, for, with respect to consumption, costs will be marginal at some times and nonmarginal at others. For example, if the safe yield of a reservoir is less than fully utilized, the only costs immediately attributable to additional consumption are certain additional operating and maintenance costs. These represent short-run marginal costs. Long-run marginal costs, on the other hand, refer to the sum of short-run marginal costs and marginal capacity costs; the latter are defined as the cost of extending capacity—for example, building a new reservoir—to accommodate an additional unit of consumption.

Now that we have two definitions of marginal cost, one applicable in the short run and the other in the long run, what happens to the rule that price should equal marginal cost? Strictly interpreted, the rule requires that price should equal short-run marginal cost when capacity is less than fully utilized, but if demand increases so that existing capacity becomes fully utilized, price should be raised to ration existing capacity. This procedure should continue up to the point where consumers reveal their willingness to pay a price equal to short-run marginal cost plus the annual equivalent\footnote{29} of marginal capacity cost. At this stage, that is, where price equals annual equivalent long-run marginal cost, investment in capacity is justified. Once the investment has been carried out, however, price should fall again to short-run marginal cost, for the only real costs (or opportunity costs, in terms of alternative benefits forgone) are then operating costs. Price, therefore, plays the roles of (a) obtaining efficient utilization of resources when operating at less than full capacity, and (b) providing a signal to invest.

Problems associated with strict marginal cost pricing, as just described, are particularly apparent in the presence of capital indivisibility, a condition typical of water supply projects, where productive capacity is often installed to meet demands for a number of years hence. Initial costs of constructing reservoirs and laying connecting mains are usually very high in relation to operating and maintenance costs. Strict marginal cost pricing in these circumstances would entail significant fluctuations in price, a source of considerable uncertainty for consumers, which would create particular problems for planning long-term investment in facilities complementary to, or competitive with, water consumption. Exploitation of groundwater—the primary source for rural systems—often gives rise to less difficulty in this respect; in the economist's jargon, the long-run marginal cost curve is frequently relatively "smooth." Even where it is technologically possible to extend capacity in fairly small increments, however, fluctuations in the availability of finance may mean that capacity is extended in large lumps. This issue is particularly important in developing countries, where large backlogs in supply may be remedied and excess capacity created at the same time.

One solution—necessarily an imperfect one—to this problem is to define marginal cost more broadly, and to set price equal to the average unit cost of incremental output. Average incremental costs can be calculated by dividing the discounted value of future supply costs by the (similarly discounted) amount of additional water to be produced. In practice, any version of marginal cost pricing has to be approximate, and ultimately some averaging of costs over a range of output is always required. Average incremental cost pricing will be theoretically less desirable the greater the degree of capital indivisibility, for while capacity remains idle, price will be in excess of the currently relevant marginal cost. However, in view of the difficulties inherent in any system requiring fluctuating prices,\footnote{30} this method appears to be the best practicable approximation to optimal pricing that can be achieved in the water supply field, and is the one that, in general, we recommend.

The characteristic of capital indivisibility is demonstrated in an extreme form by a distribution network: prior to its inception it is by definition a marginal cost and, presumably, is a function of the expected consumption of those benefiting from it. It is, how-

\footnote{29} The annual equivalent A of a lump sum expenditure E is here defined as:

\[ A = \frac{E_i (1 + i)^n}{(1 + i)^n - 1} \]

where i is the rate of interest and n is the expected useful life of the project. Any demand period could be chosen, but an annual one is obviously convenient. Note that if demand is expected to continue growing, willingness to pay a price equal to annual equivalent long-run marginal cost by consumers in the first year would imply willingness to do so over the rest of the designated useful life of the asset. If not, it simply means that the useful life has been estimated incorrectly.

\footnote{30} The political difficulties of changing prices for water supply are well known. Fluctuating prices are also unsatisfactory in that they will not provide adequate signals to consumers about long-term water supply costs and therefore will not encourage optimal investment in water using (or saving) equipment by domestic or industrial users.
ever, normally designed to meet demands placed upon it for many years hence, during which time additional consumption by existing consumers is responsible for negligible additional distribution capacity costs. The pure marginalist approach would suggest that the price charged for this element of a water undertaking's services should also be negligible. It has to be financed somehow, though, and the case is illustrative of the conflict often encountered between economic efficiency and financial requirements.

FINANCIAL VIABILITY AND ECONOMIC EFFICIENCY

Marginal cost pricing results in financial losses for an enterprise when average costs are falling, that is, when marginal cost is less than average cost. This situation could be temporary, arising, for example, where there was excess capacity and price was equated to short-run marginal cost. It might also be a situation of some permanence, even if there were perfect capital divisibility, if long-run average costs continued to decline and price were equated to long-run marginal cost. If there were lumpiness, a price equal to average incremental cost would, in these circumstances, also result in loss-making. On the other hand, if long-run average costs were rising, financial surpluses would be generated.

Any surplus resulting from the application of marginal cost pricing could conceivably be used to defray other public expenditures, or to avoid taxation, and only limited distributional or resource allocation problems would arise. Loss-making, on the other hand, may be attacked on the grounds that those who benefit should pay for a service, even though the expenditure of real resources might have taken place in the past. Indeed, the possibility that efficient resource allocation could require subsidy from those in society who do not directly benefit from the good supplied should lead the analyst to examine with care the often multiple objectives of a pricing policy. Thus, if a clear conflict of interest exists among various groups in society, the benefits of overall economic efficiency must be weighed against those of income redistribution.

Loss-making may also entail certain drawbacks from an efficiency standpoint. First, the accounting losses have to be made good somehow, and it will often be difficult to achieve the necessary transfer of real in-

31. This may not, in view of the extreme shortage of public funds in most developing countries, appear to be a problem. Politically, however, it may be, since it is often advocated that public utilities should avoid making large profits. If so, the optimal strategy might be to set price equal to marginal cost, and allow consumers to have a rebate, which does not vary according to the amount they consume.

come without creating distortions of consumer or producer choice as severe as those encountered in departing from marginal cost pricing. Second, the financial discipline and organizational autonomy resulting from financial viability are often necessary to ensure efficient operation of the undertaking concerned.

Solutions to this dilemma have been proposed that have usually tried to obtain the best of both worlds: the resource allocation advantages of marginal cost pricing on the one hand and the avoidance of financial loss-making on the other. There are, in fact, many variations on a common theme, the simplest of which is a two-part tariff where a water consumer would pay a sum per thousand gallons consumed equal to marginal cost, plus a lump sum covering non-marginal "sunk costs" and consumer costs. In this way, as long as liability to the lump sum payment does not deter anyone from consuming the system's water altogether, optimal allocation may be achieved. Similarly, efficient allocation may theoretically result from the activities of the imaginary "perfectly discriminating monopolist," who charges each consumer a price equal to the maximum the consumer would pay, right on down to the consumer who places a value on water equal to its marginal cost. Although such omniscience is rare, this general approach, popularly known as charging "what the traffic will bear," is often employed to finance water supply: for example, industrial consumers may be charged higher prices than domestic consumers. Even if these methods succeed in achieving efficiency in the short run, however, the investment decision still cannot be signalled without price fluctuations if capital indivisibility is present.

THE SECOND-BEST PROBLEM AND SHADOW PRICING

Another difficulty encountered in applying marginal cost pricing to the provision of water supplies is known as the second-best problem. What may appear at first sight to be a step in the direction of economic efficiency (for example, setting a price equal to marginal cost, or indeed, of introducing a pricing mechanism where none hitherto existed) may not be an improvement at all should inefficient conditions prevail in other sectors of the economy. Optimality in any one sector might require a price greater or less than marginal cost to counter such inefficiencies.

In practice, in any economy in which there is a good deal of competition, it has to be assumed as a rough and ready principle that elsewhere goods and services are sold at prices that in general approximate their long-run marginal costs. If not, the difficulties of adjusting for all imperfections would lead to the nihilistic conclusion that there are, after all, no empirical grounds for preferring any one set of pricing
rules over any other. Where, however, goods or services that are in direct competition with (or are complementary to) the service in question are priced in a way that diverges sharply from the standard set for the water supply or sewage disposal system, it may be necessary and feasible to make some adjustment. If prices of resources employed in constructing and operating water supplies diverge from their long-run marginal cost to society, shadow prices should ideally be placed upon them in evaluating the real cost to society of the expenditure. Thus, labor that would otherwise be unemployed might be valued near zero (that is, at its opportunity cost) even though, due to market imperfection, it is able to command a wage rate in excess of the minimum amount needed to attract it; foreign exchange costs should be valued at their natural market rate; interest rates should reflect the social opportunity cost of capital, and so on. Adjustments of this nature are necessary if the ultimate consumer is to be faced with a price for water that reflects the true economic cost his consumption entails.

Relevance of Pricing Theory for Village Water Supply

The theoretically ideal pricing and investment rules discussed above are generally somewhat removed from the methods actually used to finance and determine the value of municipal and rural water supplies in both the developed and developing world. Water authorities are normally reluctant to make use of price as a means of achieving an efficient allocation of resources, the inefficiencies generated by failure to adhere to marginal cost pricing rules being extreme when the means of implementation (that is, metering) do not exist.

The Benefits and Costs of Pricing: Existing Consumers

Where the cost to a consumer does not vary with consumption, he will continue to use water up to the point where the value to him of the last unit consumed is zero. At this point the net economic loss will be the relevant marginal cost—which, when system capacity is less than fully utilized, will equal short-run marginal cost. Inefficiency is particularly evident when, at the current level of consumption, existing capacity is on the verge of full utilization, and for some reason rationing by price is not feasible. In these circumstances, the decisionmaker is faced with the choice of permitting shortages and allocating water by nonprice means or of extending capacity. Rationing by physical or administrative means is generally accepted as being unsatisfactory as a permanent policy, although it has become the norm in many developing countries. As a public utility, a water undertaking should be able to supply water to those willing to pay for it; and as we have noted, health hazards may arise from intermittent supplies. There are also theoretical objections: nonprice rationing is a necessarily arbitrary device and can rarely be administered in accordance with the value of the benefits derived from the services rendered. It is therefore inefficient in allocating resources in the short run, and offers no guidance for the investment decision.

The policy usually preferred by decisionmakers in the water supply and sewerage sector is automatically to increase capacity when existing capacity approaches full utilization. In other words, at this point, more capacity is deemed to be required. Clearly, in the absence of a signal to invest, such as a willingness of consumers to pay a price equal to marginal cost, it can rarely be certain that the value of the additional consumption—or usage—made possible by the investment will exceed the costs thereby incurred.

Where significant budgetary constraints are lacking, a failure to employ pricing and a reliance upon the “requirements” approach will almost certainly result in overinvestment. But where guidance of the pricing mechanism is not available, budgetary constraints may prevail when, in terms of the costs and benefits of a given project, they should not. Either way, inefficiency is likely to follow.

Unfortunately, implementation of use-related pricing (that is, through metering) for water supply is a costly exercise, and its introduction or continuation should ideally be subject to cost-benefit analysis. Briefly, the benefits of metering are the cost savings brought about by reducing consumption. Savings may be achieved by deferring investment as well as by reducing annual operating and maintenance costs. To determine whether the investment in metering is worthwhile, the present worth of these savings should be compared with the present worth of initial and annual costs of metering, plus the reduction in the value of water consumed. Because the reduction in consumption likely to result from metering is normally highly conjectural, one way to approach the problem is to ask the question, what percentage reduction in consumption would justify the introduction of metering? If extreme values result from such a calculation, it is easy to make a judgment as to whether or not metering is justified; if not, at least the worst excesses of installing or not installing meters may be avoided.

The case for metering industrial consumers in urban and many rural areas is usually not a matter for serious dispute; the cost of metering is normally insignificant in relation to the cost of water consumed. The real question is whether or not to meter domestic and small commercial consumers, and here the need for some sort of cost-benefit calculation is clear. De-
spite the lack of hard empirical evidence, the water supply industry usually argues that, although metering will reduce recorded per capita consumption, changes in price, once meters have been installed, generally appear to have an insignificant effect on consumption. This apparent paradox is conventionally explained in two ways. First, universal metering reveals discrepancies between the quantity of water going into supply and that actually received on consumers' premises. Such discrepancies, revealing the extent of leakage from the mains and of illegal connections, at once facilitate and provide an incentive for the improvement by water authorities of waste prevention methods, for waste outside of registered consumers' premises would, prior to metering, have been recorded as domestic consumption. Metering would therefore bring about a permanent reduction in annual wastage, a condition that would obviously remain unaffected by subsequent price changes. Second is the argument that metering will encourage individuals to reduce the amount of water wasted on their premises, but once having made this adjustment, their demand for water remains relatively inelastic. Intuitively both arguments would seem to be valid.

There are a number of special problems, reflected in both the cost and benefit sides of the calculation, which suggest that metering will rarely be appropriate for residential and commercial consumers in rural areas. The costs of metering will tend to be relatively high, for meter reading and maintenance will be more costly because of lower population density (particularly a population wealthy enough to have household connections). Furthermore, in smaller communities the meter reader may be under greater social pressure to underrecord consumption. On the benefit side, consumption tends to be lower in rural areas where people are poorer and do not have water-using appliances or adequate means of disposing of wastewater. The potential gains from conservation of water are therefore limited. Exceptions of course exist; the lower density of population may mean that more water is used for garden watering or for livestock; where this is so, metering is a more viable proposition.

Central to the problem of metering, however, is that, because of low incomes in rural areas, the appropriate source of supply may not be household connections but rather communal standposts or wells with hand pumps. Although in these circumstances metering, in the normal sense of the term, may not be an appropriate means of influencing consumption, principles similar to those outlined above should be applied to the decision to use a village or water board employee to supervise the dispensing of water from standposts, or to sell the concession to an individual. These methods are frequently used in urban areas of developing countries where standpost supplies are provided, although somewhat less so in rural areas. Where they are used, selection of the means of control is sometimes based on financial considerations, rather than on a comparison of economic costs and benefits, as described above. This is not always so, however, for conservation of water and the prevention of vandalism to taps—which presumably should be defined in cost-benefit terms—are often cited as the reason for employee control of taps or sale of concessions.

Kenya provides an interesting example of a case in which water vendors (water kiosk operators) in a number of rural villages serve the multiple functions of generating a small amount of revenue from public taps, of limiting public tap water wastage, and of protecting the taps from vandalism. In several regions of the country, villagers had not been paying the small monthly tax which was to be used to help operate and maintain their local water supply systems. Furthermore, it was financially very costly and physically almost impossible to maintain many of the public standposts because of frequent acts of vandalism on taps, drainage facilities, protective fences, and so on. As a result, in a few areas the public standposts were converted to water vendor operations in which a licensed vendor is paid a subsidized rate for the metered standpost water and sells it by the container (debe) at a fee slightly higher than he pays. The difference between the kiosk operator's buying and selling price does not have to be large because most of the rural kiosks are operated by the wife or children of the vendor, and their opportunity cost is low. The result of the switch to kiosks is that vandalism has been greatly reduced (the tap and meter are locked up when not in use), a small amount of revenue has been generated, and the rate at which people have applied for house connections has increased. (Some people presumably feel that, if they are going to have to pay for water, it might as well be convenient water.) A drawback to the system, however, is the inability of some kiosk operators to retain for any period of time the revenues collected from water sales, thereby forcing the water authority to collect its sales revenues daily from the kiosk operators, a rather costly alternative.

NEW CONSUMERS OR COMMUNITIES

The previous section dealt with consumers who currently have a supply of water, and who may or may not be charged on the basis of the amount they actually consume. The question of pricing as an indicator of project benefit takes on a different complexion when related to the problem of deciding whether individual households or communities, hitherto relying on private sources of supply, should be
supplied from a public system. Here, by definition, direct ex ante observation of the willingness of individual consumers to pay is not possible. We would also argue that the questionnaire approach to estimating individuals' willingness to pay has been shown to be virtually useless.

As far as new communities are concerned, the approach to this problem should be twofold. First, as an example, in India and in many Latin American countries, the necessary qualification for supply should be a community contribution, say of 20 to 30 percent of initial project costs. Where the smaller villages are concerned, this contribution may be a reasonable indication of the value placed upon the project; the larger or less democratic the village, and the more activities that it carries on as a community, the less this proxy can be relied upon.

This screening device should be supplemented, as far as possible, by analysis of other communities in the country in terms of the observed willingness of their inhabitants to pay for water (clearly such analysis should be extended to administrative and technical aspects). The greater the variation in the economic, cultural, and climatic conditions of the villages concerned, and the smaller the sample of such villages, the harder will be the task of estimating willingness to pay in those areas as yet unserved. Even in these circumstances, however, this discipline is valuable for policymakers in determining investment priorities, and there is no real alternative to the procedure.

A somewhat similar approach should be followed in deciding on whether or not to extend service to new consumers in areas currently served. As far as household connections are concerned, there is no real problem—individual willingness to pay for connection can easily be demonstrated. Where the issue is whether or not to extend standposts into a new area of a community, there would be no problem if (a) the consumers concerned had cultural and economic characteristics fairly similar to those in other parts of the community (or in other communities) who currently had standpost supplies; and (b) if those currently served were able to demonstrate willingness to pay by the means described in the foregoing section. Where these conditions do not exist, little guidance is available to the policymaker as to the economic merits of the investment, and a method of trial and error is his only course of action.

THE PROBLEM OF GEOGRAPHICALLY UNIFORM TARIFFS

A particularly important application of the incremental cost-pricing principle is to be found in the establishment of a charging policy for nationwide or statewide water boards, which may serve hundreds of different communities. The question at issue, one highly relevant for village water supplies, is how one should respond to what seems to be a mounting pressure for geographically uniform prices. This pressure is important, not only because of the strength of the emotional appeals for uniform water rates, but also because of the wide variations in costs of supply within relatively small geographical areas.

The observed tendency toward uniformity results in part from the growing recognition that, in the interest of operational efficiency, particularly where skilled manpower is at a premium, consolidation of the management of water authorities into large regional, state, or countrywide water boards is often desirable. For no sound reason—except to some extent in those cases in which the water board is set up to permit a grid system to be established where at least marginal production costs will tend to greater geographical uniformity—the replacement of a number of small water authorities by one water board often involves the replacement of many different pricing policies by one uniform tariff schedule.

Another explanation for the trend toward increasing uniformity is the improvement in communications which allows residents of remote communities to become aware of the prices that consumers in other parts of the country are being asked to pay for water. Parallel to those arguments which have been heard in the past regarding variations in water charges within cities and which have been successful in eliminating surcharges for consumers who live on hillsides or in the less densely populated suburbs, it is alleged that only a uniform rate of charge is equitable. Because, it is said, a gallon of water is a gallon of water wherever it is supplied, a village should not be penalized if supplying it with water happens to be relatively expensive.

There are many examples one can cite of the unfortunate effects of uniform water pricing on land use. In one African country in which uniform water charges are in force, a brewery was located in an area where existing water supplies were of a sufficiently high quality but capacity was on the verge of full utilization. Although additional water could readily have been supplied, its extremely high fluoride content meant that it could not be used for brewing beer, and treating it would have been very costly. Convenient sources of low fluoride water soon ran out, and the brewery eventually closed down, to be reopened at a later date only when high-cost water was transported from many miles away. The initial brewery-location decision in this case was most inefficient, and stemmed in effect from a failure to levy charges for water (of an acceptable quality) that reflected incremental costs.

Aside from their direct impact on locational choice, if prices reflected the incremental costs of supply in different areas, a valuable discipline might be im-
posed on regional or urban planners. In the same African country, a number of development areas, or "growth poles," have been designated. As uniform water charges are levied throughout the country, the choice of industrial location is not influenced by the cost of water. Unfortunately, the incremental costs of water for many of the growth poles are not known even by the planners. If industry expects to be faced with costs which vary according to geographical location, it would make inquiries as to expected water or other charges, thereby forcing planners to carry out the necessary cost estimates, and indirectly to improve the efficiency of their plans.

Paradoxically, the geographically uniform water rate is also frequently the cause of further adjustments to tariff structures that are unwarranted by variations in cost. One argument used in developing countries to defend the uniform rate is the necessity to ensure that the poor obtain water, based on the assumptions, first, that economies of scale and of population density mean that urban areas are the cheapest to supply on a per capita basis; and, second, that the smaller the community, the poorer the people. Whereas, for a given standard of service, the first assumption will usually be true, there are, as has been previously discussed, some problems in blindly accepting the second assumption. Given the goal of providing cheap water to poor people, the geographically uniform tariff is usually associated with an increasing block rate in which the more one consumes, the higher the price paid for incremental consumption. In efficiency terms this might produce the worst of both worlds: between villages, prices are uniform when costs vary; within villages, there are variations in price which do not reflect cost differences. Moreover, quite apart from the allocative inefficiency of the increasing block rate, it may also have adverse income distributional consequences.

Although the adjustment is designed to take advantage of the presumed positive income elasticity of demand for water, the final outcome may still be the effective subsidization of relatively high-income rural consumers by low-income urban slum dwellers. Moreover, even within an urban or village area, the increasing block rate may be regressive in its effect, a fact illustrated by another brewery example—this time in a small South American community. As the largest water consumer in the community, the brewery also paid the highest per unit price for water, much higher than even the wealthiest residential consumer. It is probable that the higher price for water was reflected in higher beer prices, and because beer formed a relatively high proportion of the expenditure of low-income groups in the village, this policy was probably regressive.

If, where variations in price are administratively feasible, uniform pricing is chosen instead to achieve certain objectives, such as maximization of the number of consumers or extension of supply to low-income groups, a form of price discrimination is in effect being practiced. In common with most other discriminatory pricing schemes, this approach is likely to yield an inefficient allocation of resources and to result in inefficient locational decisions. Indeed, the uniform-charges policy is such a cumbersome instrument that it will only be by chance that the intended objectives are in fact realized. It is a perfect illustration of the difficulties which arise from not using cost as a basis for establishing water tariffs.

OTHER METHODS OF REGULATING CONSUMPTION

Where direct regulation of consumption by price is ruled out, a variety of methods may still be called upon by water authorities to conserve water. Some of these allow consumers' willingness to pay for water to be revealed, albeit indirectly, and therefore have the characteristic of a pricing policy. Others rely on legal or social pressure, or on physical restrictions. Some examples follow.

Individual house reservoirs. The rural water supply program in Argentina makes use of a system of individual house water tanks, or reservoirs, both to regulate the flow of water and to charge according to the volume of water consumed. A cement tank sitting on top of each house has a small hole through which the water flows, taking a considerable time to fill up the tank. The inhabitants of the house can use the water in whatever quantity they want until the tank is empty. Generally, the water users ration their consumption over the period of a day, so that the tank is never quite dry except perhaps in the evening. The tank can then refill throughout the night. People pay different water charges according to whether they want tank inlets (orifices) rated at 300, 500, 800, or 1,000 liters. The tanks, or cement house reservoirs, come in two sizes: 300 and 500 liters. The 300-liter tank serves the 300- and 500-liter inlets (the orifice is larger so that the tank refills more rapidly for a 500-liter inlet); the 500-liter tank serves the 800- and 1,000-liter inlets. The price charged for water varies by village but, in general, the monthly water rate for the 300-liter inlet is approximately 1 to 2 percent of the monthly income of the water consumers. This device therefore acts as a means of pricing and regulation of consumption and is probably much cheaper than metering.

The one potential danger of a system like this, however, is that, since all water consumers have roof tanks, there may be the temptation at some future date to provide users with intermittent service. Such an event could result in a marked deterioration in the health benefits of the system.
Flat rates, with rules about consumption. In many countries a flat monthly rate is charged for each house connection, with rules about how the water may be used. The rules sometimes specify that only the family cow, horse, and chickens may be watered with the potable water, or that certain plants or trees that consume a lot of water cannot be grown near the water tap. Often the wasting of water is prohibited (failing to turn the tap off), and the village water committee, or council, has the right to disconnect people's water supply either for water wastage or non-payment of water bills.

In areas where some of the residents desire to water their livestock during dry seasons, there may be communal animal-watering stations, or residents may be allowed to purchase an extra water outlet and have it installed near their residence so that they can water their livestock there. In the case of communal livestock-watering stations, water might be provided free or each resident might have to pay a monthly fee for the right to water livestock. Where the resident acquires an additional outlet to water his livestock he usually has to pay a rate approximately equal to what he would have to pay for another individual house connection. In countries that have financially viable water supply programs, and that charge for water, the monthly flat fee water rate for house connections rarely exceeds 5 percent of estimated monthly income. Several forms of price discrimination, in addition to that based solely on quantity of water consumed, are also generally practiced. For example, in a number of countries most rural household users are charged a flat rate, whereas any large users in the village (village industry, and so on) are metered. Also the use of outlets for livestock watering, or for gardens, can be charged at a higher or lower rate than the household rate. In some areas of Kenya, for example, the domestic rate is set in an attempt to cover both fixed and variable costs whereas the cattle rate is supposed to cover only variable pumping costs.

In most countries, different rates are charged to families that use public fountains, or standposts, than to those that have house connections. Those with house connections have to pay not only an initial connection fee but also a higher monthly charge than those who use public fountains. In many instances, despite the greater cost of house connections, the proportion of the village with house connections increases fairly rapidly through time. Real incomes increase and there seems to be a demonstration effect existing: as people begin to perceive the advantages of having a house connection they somehow manage to save the money necessary to acquire and maintain one. The trend is sometimes reinforced (as in several Latin American countries) by the closure of public fountains after a certain percentage of the village has house connections, leaving the remainder little alternative but to connect also, or to purchase water from neighbors.

In some countries and cultures the collection of monthly flat-rate water charges is difficult. Generally, arrears seem to be less of a problem in Central and South America than in some portions of Africa where water is viewed as a special gift from God and therefore always free. Head taxes or house taxes for water in some rural areas of both East Africa and West Africa have been quite unsatisfactory, with collection rates of less than 20 percent.

Social pressure and physical restrictions. When a flat-rate water fee is instituted, water wastage can, in some countries and cultures, be a problem. There are essentially two means of dealing with this problem: community moral suasion and the installation of flow-limiting devices. Throughout much of Latin America and in some areas in Asia, country officials claim that community moral suasion has been sufficient to keep water wastage within acceptable limits.

The Willingness- and Ability-to-Pay Criterion

There are a number of arguments against relying entirely upon the willingness of consumers to pay as a criterion for supplying them with water. These include consideration of external benefits, the extent of consumer knowledge, and the ability to pay.

EXTERNALITIES, CONSUMER KNOWLEDGE, AND ABILITY TO PAY

An external benefit that might result from the consumption of potable water is that the health of X might improve because he makes use of an improved supply, and as a consequence X may not infect Y, whose future health would also improve. However, since X would not take the health of Y into account in his decision to consume potable water, his willingness to pay would tend to understimate the benefits that would accrue to the community as a whole. Another argument is simply that the villagers may not understand (like the rest of us) the relationship between improved water supply and health: the assumption of a well-informed consumer is essential if normative judgments are to be made about the expression of his willingness to pay.

Apart from the issues of externality and consumer information, both of which are primarily relevant to the health question, the major criticism of the willingness-to-pay criterion is that, if strictly enforced, it would effectively mean that large numbers of people would never, or at least would not for many years, obtain adequate water supplies. To be a useful concept, willingness to pay has to assume an ability to pay. As noted earlier, the combination of low per
capita incomes and higher per capita costs (for a given quality of service) in rural areas constitutes a formidable obstacle to improvement in the rural water supply field. Any rapid expansion of service to the rural poor requires that the willingness-to-pay criterion be modified to allow low-income groups to obtain at least a basic supply of water for minimum health needs. In practice, therefore, subsidy might be necessary, and current practice in this regard is described below.

**FINANCIAL SUBSIDIES AND ABILITY TO PAY:**

**CURRENT PRACTICE**

We have observed that many different methods of charging for water are used in developing countries, but in most countries it is fairly clear that rural consumers normally pay less than average system costs, and frequently even fail to cover operation and maintenance costs.

Since the primary rationale for subsidization of rural water supply is the lack of the beneficiaries' ability to pay, two questions should be investigated: the average real income of the population and the amount of that income that can be spent on water. Money income may be very different from real income in rural areas, primarily because of the prevalence of subsistence agriculture. With little or no reliable data available on either money or real income in rural villages, the estimation of the real income of rural populations is an extremely difficult task.

Even if the real income level was known there is the second question: What amount of a family's real income can be spent on water? A frequently used rule-of-thumb reply is that a rural near-subsistence family should never have to spend more than about 5 percent of its income for water. This 5 percent-of-income figure is usually more than most urban dwellers pay for the water they consume from the public system.

There are several reasons, however, why any reliance on an arbitrary 5 percent of income rule as an "appropriate maximum percentage" is fraught with hazards. Clearly, the "appropriate percentage" will be determined in large measure by the degree of monetization of the local economy, the cost of other essential items, and so on. Furthermore, who is to decide what the word "appropriate" means in each case? Who decides that the villager should give up one pair of shoes each year in order to pay a monthly water charge that covers total cost, or that is equal to marginal cost?

Even if a villager can pay for water, he may not be willing to alter his current expenditure pattern, or reduce his slight savings each month, in order to improve the quantity or quality of water he and his family consume. If this is the case, there are two alternatives available to the water supply authorities once they have decided that the villager needs, and will be provided with, an improved water supply. First, they can attempt to alter the villager's perception of the "value" to him and his family of a better water supply. There are various means by which this might be done, but generally they involve some form of demonstration or education program acquainting the villager with the more obvious benefits of improved water and sanitation. If this education function is successful, the villager will then perceive greater benefits from the improved water supply or sanitation system, value the improved system more highly, and therefore be willing to pay a higher price for it.

The second alternative is for the water supply authorities to subsidize the village system, using the rationale that the unrealized or unperceived benefits which accrue to the villagers, together with the benefits external to the villagers but which accrue to society as a whole, make the subsidy worthwhile.

Given the difficulties described, we have particular doubts as to the practical relevance of the concept of ability to pay in arriving at a judgment about willingness to do so. There is much evidence that even where consumers are wealthy, they often refuse to pay or otherwise cause difficulties for a water authority which is attempting to introduce or increase water charges. This leads us to suggest that the only way in practice to address this issue is, rather than to conduct elaborate, assumptive socioeconomic surveys, to "test the market" by the gradual introduction of new tariff policies, and then to observe consumer reaction before deciding to increase those charges or to expand capacity.

The general lack of any hard evidence on ability (and willingness) to pay has resulted in the politically expedient assumption, which has been made in most developing countries, that the rural population cannot pay the full cost of water. As noted earlier, in a number of Asian and Latin American countries, the beneficiary village is expected to make only a partial contribution of between 10 and 30 percent to construction costs, and to pay only for the operation and maintenance of the system. Experience in these countries suggests that this is a minimum prerequisite for program viability. In a few countries, however, rural consumers do not pay at all. These countries argue that potable water is a basic human right, irrespective of ability to pay, and as a result the national government should completely subsidize the rural water supply program from national general revenues. Tanzania and Zambia provide good examples of this philosophy, although their neighbor, Kenya, does not generally subscribe to it.

The most prominent or frequently stated rule of thumb about an acceptable minimum a village should be required to pay to be eligible to receive an im-
proved water supply system in the first place is that the village should at least be able to cover ongoing operation and maintenance expenses. This principle probably has as its base a form of administrative rationality in that it simplifies local accounting, and it bestows upon the villagers the psychological satisfaction of paying for what they can see is taking place. In addition, there is the financial consideration that capital funds are relatively easy to budget, get grants for, and so on, whereas adequate funds for operation and maintenance tend to be more difficult to generate from central government sources. From an economic point of view, however, if the villagers are to be partially subsidized, it makes no difference whether the subsidy occurs for their operating expenses or for the capital costs of the project. A capital costs–operating costs dichotomy for purposes of subsidy can hardly affect the magnitude of the health and economic benefits the villagers receive from the project.

Cost Recovery

Another major problem arises if marginal cost pricing does not generate sufficient revenue to cover all the financial costs of a project. If average total cost remains greater than price because of decreasing costs and economies of scale, the problem of cost recovery must be solved in some other way. Beyond the efficiency dimension of pricing policy, there must then be consideration of the financial dimension.

There are essentially three cost recovery "theories": (1) theories based on benefit measurement, according to amount of use or to differential benefits; (2) theories based on cost causation; and (3) theories based on economic efficiency—namely, social marginal cost pricing. The next reading examines this problem of covering financial costs and the relations between the financial and efficiency dimensions of pricing.

4.11 Public Transport Pricing and Cost Recovery

As a preliminary to subsequent discussion, it is necessary to set out the general principles of public enterprise pricing. It is usual to start by taking a period of time of such length that no question of replacement of capital equipment arises. It is also convenient to assume initially that one is concerned only with minor adjustments of output, and that such questions as that of closing down altogether rather than continuing in production or that of making major extensions to capacity do not arise.

In these circumstances, the conventional argument on public enterprise pricing runs as follows (all references are to figure 4-4). There are a number of ways in which one can relate costs and output, but the first customary distinction is between variable costs and fixed costs (i.e., between those costs whose total varies with changes in output and those whose total does not). The latter are shown by AFC, a rectangular hyperbola; the former by MVC and AVC, MVC showing marginal variable costs and AVC average variable costs. The precise shape of AVC is a matter for some argument, but it is often maintained that it will tend to fall as output increases from a very low level (e.g., through more intensive usage.

Figure 4-4. Public Enterprise Pricing

of some indivisible input) and to rise again once output expands to a high level (due to, e.g., excessive strain on plant). But, whatever the precise shape of AVC and MVC, we shall find that the latter falls and rises faster than the former. AGC (average global costs) is derived by vertical summation of AVC and AFC. MVC will intersect AVC and AGC at their lowest points.

Leaving aside for the moment LRMC and LRAC, the first proposition is that any public enterprise should equate prices of products to MVC. Only in this case will the marginal value of output to consumers be equal to the alternative use value of the resources engaged. If price is set at a lower level, output will be greater than optimal and too many resources will be engaged in that activity; if price is set at a higher level, the converse will hold. In applying this principle, it is necessary to take account of (the present value of) any costs consequentially incurred at later dates during the planning period.

Several more points should be noted about this proposition. First, it only holds fully for any one industry on the assumption that all other industrialists operate on similar principles, both in selling goods and in hiring factors. Second, it is assumed at this stage that various side effects such as third party effects ("externalities") or consequences of calling on general government revenues to finance deficits can be ignored. Third, strict adherence to the principle can lead to great volatility in prices if demand suddenly changes or if there are major discontinuities in variable costs. Even the strictest devotees of the principle usually temper it by saying that averaging is needed either for administrative reasons or because the time taken by consumers to react may be longer than the intervals between price changes which strict adherence to the rule would dictate.

A further point to note—and this applies to other versions of marginal cost pricing as well as the very simple one we are dealing with at the moment—is the relationship to the benefit principle familiar in the literature of public finance. Confusion is liable to arise in this matter because public utility pricing systems based on the maximum which the traffic will bear are sometimes referred to as "charging according to benefits." The distinction here is simply that between total and marginal benefits. Whereas a system of price discrimination approximates in the limit to one of charging according to total benefit (in that total revenue is then equal to the whole area under the demand curve and consumers are deprived of all consumers' surpluses), a system of marginal cost pricing secures the equality of benefits with costs at the margin (in that if price, taken as equal to marginal benefit, exceeds marginal cost there will be an incentive to expand output until any difference is eliminated).

Finally, we have to ask about the shape of the appropriate demand curve. Harrod has argued strongly that most firms do, and on balance should, look at long-run rather than short-run demand curves—the former being more elastic at more points than the latter—even in taking short-run pricing decisions. The same principle will surely apply to public enterprises. They are usually subject to some kind of government regulation on prices and as such procedures are usually lengthy one cannot initiate them all that often. Furthermore, they tend to be conscious of a public image of respectability, conformity to established norms and practices, permanence, and so on. If the relevance of a long-run demand curve is accepted, then the need for observance of the standard pricing rule is all the greater in that, with greater elasticity of demand than in the case of short-run curves, the greater will be the impairment of resource allocation in the case of departures from the rule.

Transport Applications

So much for very general principles. What do they entail in the case of transport? The first and most general point is that they imply that prices should be fixed on the basis of marginal operating costs and that no notice should be taken of fixed costs. So in railroad pricing the relevant question is what incremental costs are incurred by conveying, e.g., an extra ton of coal; the interest charges on existing debt are irrelevant. Similarly, with roads, one has to ask what additional maintenance costs result from their usage by an extra truck or bus. At the same time, discontinuities are particularly obvious—the classic example is that of the extra railway passenger who could be responsible for anything from zero additional cost, if he could find an empty seat in a train which was running anyway, to the provision of an extra coach or even an extra train. This ties up with another point. We emphasised that the simple argument about marginal variable costs assumes a given stock of capital. When the capital stock consists of a large number of items rather than a single one, this notion becomes very complex in that if we only allow a very short time to elapse none of the stock will be variable whereas a slightly longer time will allow some to be variable but not the remainder. Even the notion of length of time is not really the fundamental issue; the really crucial point is that of expenditure commitments or contracts which have already been undertaken. The longer the time period one looks ahead the smaller is likely to be the element of commitment and the larger that of variability; but there is clearly no simple relationship between length of time and degree of commitment—it may vary for the same
enterprise at different points in time or between different enterprises at the same point in time.

There are many more complications. To illustrate, rather than to treat exhaustively, one must take into account any effects of decisions about pricing of current outputs on both the costs of and demand for future output within the given time horizon. In addition to these issues of temporal linkages, there are also those of system linkages. For instance, it may be sensible to price the usage of a rail or road feeder service below marginal cost if a higher figure would mean a reduction in usage of a main line or main trunk route and hence a smaller incremental return for the system as a whole. Similarly, it is not necessary that prices should cover the marginal costs of transport from A to B and those from B to A separately; insofar as the two journeys can be regarded as a joint product, the only necessary requirement is that the price for the two together should cover the marginal cost of the joint operation.

The problem of cost definition is intensified by the difficulty of defining a standard quality in a service industry like transport. Pinning down the marginal variable costs of providing road space for an additional lorry depends on one's ability to say, inter alia, when the standard of a road's surface is kept constant; and the marginal variable costs of carrying an extra passenger on the railway might appear to be constant but would in effect be rising sharply if the discomfort of travel rose sharply due to more passengers being crowded together.

Despite these difficulties of interpretation of the general principle, it has now been argued for a long time that, by and large, transport pricing should be based on it.

It is easy to see why people have felt this way. If we predicate a long-run demand curve, the possible wastes of not conforming to the marginal variable cost principle become especially high, as we not only have the more obvious failure to use existing capacity to the best advantage but also the point that decisions about the location of plants and factories may be adversely affected. However, elasticity of demand depends, as always, on the degree of substitutability between products and so the chances of undesirable repercussions are less if one is looking at pricing policy for the whole transport system rather than at a single means of conveyance between two particular points. They are also less if public transport charges are only a small part of total costs.

Finally, one must be clear that the assumption of marginal variable cost being equal to price throughout the rest of the economy is most unlikely to be valid. There is no short cut prescription for amending the formula in such cases, but we shall return to some of the complications later.

The Closing-Down Case

Our analysis so far has been confined to marginal costs on their own but now we must consider them in relation to other cost concepts. First, what are the consequences when marginal variable costs are less than average variable costs, AVC, as in the case of low levels of output in figure 4-4? It is often argued that if the demand situation is such that pricing on a MVC basis would not cover AVC, the firm should cease operations. All due allowance must, of course, be made for the repercussions of such a move—expected costs of restarting if there is any possibility of demand improving, loss of customers if supply is intermittent, etc. Subject to such modifications, the portion of the AVC curve to the left of A, the point of intersection with MVC, can be thought of as being the appropriate marginal curve relevant to decisions about whether to produce or not—in the sense that it shows the savings which can be made by ceasing operations altogether.

Two major comments have to be made on this notion. First, if there are non-marginal changes in the process of shutting down, the fundamental criterion on which the decision to produce, or not to produce, should be taken is not the comparison between changes in total revenue and in total costs,
but rather that between changes in total utility and in total costs. This is a well-known point going back (at least) to Marshall’s *Principles* and reiterated on a number of occasions. If the position is as shown in figure 4-5 where OBC shows the total utility from producing OC and OEGC the total cost, then production is preferable to closure (even though the demand curve lies below AVC throughout) provided that AFB exceeds AEGB in size. The essential point in a case of this sort is that if it were possible to discriminate perfectly between consumers, total revenue could then be greater than total costs. For the moment, we shall leave on one side the practical importance of this sort of case in the transport field and return to it later.

The second major comment is that decisions about closing down or opening up production may be related to those about capital investment. In some cases, the decision whether to close down or continue may have no implications for the state or condition of an undertaking’s capital equipment; but it is also possible to envisage a situation where capital equipment is more or less worn out and where the decision whether or not to continue production is indissolubly linked with whether or not to replace capital equipment. This is one facet of the old tag that in the long run all costs are prime costs.

**Other Reasons for Deficits**

Returning to figure 4-4, we have now dealt with some of the issues arising when MVC < AVC. We now look at other points on the MVC curve and start with B. If the demand curve intersects the cost curve here, it is a point of both short- and long-run equilibrium; short-run in the sense that output at that level satisfies the MVC = Price condition and long-run in the sense that not only is the sum of average fixed (AFC) and average variable (AVC) costs ( = average global costs, AGC), covered but also are long-run average and long-run marginal costs. If the demand curve does in fact pass through B there will in general be a case for renewing capital equipment of the same capacity, neither more nor less, whenever necessary.

If the demand curve, however, cuts MVC at a point to the left of B, we then come to the general problem of public enterprise deficits—in that if output is determined on the MC = Price principle, average revenue must be less than average global costs. There is one point to note before we face up to this issue, however: if the decision-making period is so long that there are no fixed commitments, we then have a state of affairs where all costs are variable. In that case we no longer have a divergence between AVC and AGC (and also LRAC and LRMC become superfluous). We are therefore left with only MVC and AVC (both drawn on a new scale) and no problems of losses arise (unless demand is so small that price lies below AVC). In that case, we are back to the same prescriptions as in the short-term variable cost case. Leaving that possibility on one side, it will be apparent that pricing on the basis of marginal variable cost will lead to a deficit. Before considering the finance of this deficit, let us look at the implications for replacement of the capital stock. Obviously, the normal expectation would be that if losses have been incurred, then at the very least there will be a contraction of the capital stock when the time comes for renewal, or, perhaps, even a complete shutdown, depending on expected future demand. But once we contemplate large changes in output of this sort, we have to resort to the total benefit-total cost principle. In other words, whilst a calculation in terms of average revenue and average cost might give one size of plant when investment decisions are taken, a contemplation of average utility and average cost might give us a larger one. In the latter case, the result will be a persisting deficit which will have to be met by one means or another.

The standard illustration of this kind is when LRMC is downward sloping; this is a highly important and relevant case in many transport situations (e.g., the provision of mass transit). Figure 4-6 shows the position.

Suppose we start from P1, the point of intersection between the demand curve on the one hand and MVC1 and LRMC on the other. Price would be less than AGC1 and so a deficit would be incurred. One
would then expect to see plant size and output contracted, say, to $P_z$, the point at which $D$ intersects MVC. This would now mean that average global costs $AGC_z$ are covered. But if we view the change as one which must be decided on a utility-cost basis rather than a revenue-cost basis, it is then perfectly possible that the decision to contract cannot be justified. We have therefore a case in which a deficit persists even though it could be avoided. In other words, what started as a position of unavoidable loss ends as one of avoidable loss. The principles on which such a deficit should be met will be explored shortly but in the meantime it should be noted that it is not the presence of scale economics as such which makes the deficit persist, but the application of utility rather than revenue considerations.

**The Finance of Deficits**

It would appear therefore that there are two distinct reasons why a public enterprise operating on a MVC basis may be running at a loss at any one time. First, there may be an unintentional loss—the potential demand over-estimated, capital equipment too big, etc. Second, there may be an intentional one—in the sense that total benefits or utility calculations predicate a larger capital equipment than will allow a surplus, given the usual MVC pricing principle. In practice, it may be extremely difficult to distinguish the two components (e.g., an intentional deficit may unintentionally increase in size) but analytically it is most important to do so.

A deficit of the first type must be due to an unforeseen shift of the demand curve or of the marginal variable cost curve, or both, making some capacity redundant. In these circumstances, the natural thing would be to say that the penalty of the mistake should fall on producers, by analogy with what would happen in a privately owned industry. In that case it may be possible to draw on reserves for a limited period, but sooner or later stockholders will feel the draught. But as public utilities frequently do not have equity stock, there may be no equity shareholders to take part or all of the strain. Nor does one expect to find such categories as holders of preference stocks, who can be squeezed if a mistake of this sort occurs in private industry. Indeed, it is precisely the fact that public utility stocks are deemed to be free of that sort of risk which allows them to be issued at lower coupon rates than large industrial issues of similar standing and maturity. The upshot is that one is forced back to government subsidies of one sort or another, until such time as the industry has contracted sufficiently to obviate the need for them. However, there are at least three reasons why such a financial solution may be undesirable. First, the subsidy from public funds can only come from more public revenue, more borrowing or less spending; it is all too easy to forget that the consequences of any one of these courses of action for the supply of total resources of their allocation between different ends may be substantially worse than the loss of benefits to consumers through paying a price higher than MVC for the consumption of transport services. Second, it may be judged that on income distribution grounds there is a stronger case for meeting the deficit out of the pockets of transport consumers rather than from those who would pay the higher taxes or receive less benefits from government spending. This might be the case if, for instance, the only possible immediate way of raising more government revenue was excise taxes on articles of mass consumption. Third, there are various administrative reasons why one may not wish to go the whole hog. There is the well-known argument that, once started, government subsidies breed expectations of further subsidies and so are conducive to inefficiency in management. There is also the danger that organised labour pitches its demands for wage increases that much higher if it thinks that the public purse can be drawn on for this purpose.

For reasons of this kind, one may wish to saddle the particular consumers of a service with some of the losses due to unintentional over-equipment. What about the case of intentional over-equipment? Here the answer would seem to depend on whether the excess of total utility over total cost is derived solely from considering benefits to the current generation of consumers or whether their successors' welfare is also being taken into account. Once again this distinction may be difficult to quantify in practice but at least we must discuss the principle.

If we are only concerned with the present generation of consumers, it would seem reasonable to argue that charges on them should, by one means or another, be high enough to prevent a deficit emerging. If the deficit is due to selecting capacity size on the criterion of total utility exceeding total costs, then by definition it is the consumers of that service who are gaining from this decision and it is hard to see why the burden of deficit finance should be borne by taxpayers at large. There is endless scope for argument about the best way of covering such costs. Discriminatory charges are one method—and allegedly a very traditional one in railway rate making. However, there are both economic and political limits to this method. Even when discrimination is easy between different consumer groups and/or different units of consumption it may be unwise to make it severe; and practices which border on being inquisitorial (and this may be the only means of making discrimination work effectively in some cases) are liable to be resented and to be attacked through political or other channels. Another method is the
two-part tariff type of arrangement, whereby each consumer pays a fixed or quasi-fixed sum as well as according to the amount consumed. The major snag about such a device is that 'entry charges' of this sort are likely to keep out some consumers and so reduce consumption below what it would be purely on marginal cost considerations and so, in turn, reduce the usage of capacity to less than the optimum. A straight per capita standing charge might also be attacked as being a regressive tax, but this difficulty can sometimes be avoided by relating the standing charge to income. A third device is to say that consumers should be charged on the basis of LRMC and not MVC. This will automatically secure that some account is taken of capital charges as well as variable costs. But deficits are not necessarily eliminated thereby; and this principle is again subject to the attack that it may cut usage of existing resources to an unwarranted extent.

These are the sorts of devices which can be used to make consumers pay for an intentional deficit—or for that matter, an unintentional one, if it is judged that the burden in such a case should lie on consumers rather than taxpayers. All of them are practicable, provided one is willing to make approximations, sometimes pretty large ones. All are capable of application to transport, e.g., a railway might discriminate, or it might construct a fare structure on the basis of a station charge plus a mileage charge, or it may load short-term marginal cost with a capital charges factor. But there are no unique principles for choosing between them in that everyone will have different ideas about the relative merits of a slightly differing income distribution and a slightly differing usage of existing capital equipment. And all of them may be weak instruments when benefits are widespread and diffused rather than concentrated on a small and discernible body of consumers.

Finally, what if a larger capacity than can currently be operated without a deficit is installed in order to cater for the welfare of a future generation as well as the present one? This is a common phenomenon, e.g., a road may be built to a certain specification not just on account of traffic at the moment but with an eye to what it will be a few years hence; or a breakwater may have to be built to cover a whole port area even though the initial number of berths and amount of traffic are insufficient for profitable operation. We shall come to these points in more detail later; at this stage we are only inquiring about the finance of deficits incurred for such reasons. In this case, it does not seem reasonable to argue that the current deficit should be met entirely by the present set of users. There is surely a case for some finance from general government revenue. This is an example of a collective decision which involves the sacrifice of some current consumption in order to build capital equipment which will augment future consumption and so should be financed in the same way as other such decisions. In practice, there are inevitable difficulties of separating one generation from another—some people must always straddle both—but the principle is clear.

There is a further point about the provision of finance from general funds for such purposes; if one accepts any of the arguments that the burden of debt rests on future generations, there would be a case for saying that government funds should be raised by debt issue rather than by taxation in that the burden would then fall on the beneficiaries from the project. We should then have the following set-up: the element of benefit to future generations should be financed by government debt-raising and the element of benefit to the present generation should be financed by user charges (unless there are strong arguments to the contrary in the case of the latter).

Subsidies

For a variety of purposes, a government may introduce a subsidy on a commodity or service by making a payment to the producer, or by the remission of charges, or by supplying the commodity or service at less than cost or market price. A subsidy may be introduced to stimulate development or consumption of some desired activity or service, to stimulate development of a particular area, to redress deficiencies in income distribution, to reduce risks of speculative activity such as research and development, or to encourage an activity that yields external economies. A common argument against imposing a user charge, even if it is technically possible to do so, is that there should be a redistribution of income in the form of a particular service or that the service qualifies as a merit good that should be consumed in greater amounts than would result if a price were charged.

Although reasons for a subsidy may be readily presented, it is essential to evaluate whether a subsidy is the best way to achieve the desired objective, to
trace through the ultimate real effects of the subsidy, and to assess the appropriate amount of subsidy. It is also necessary to meet criticisms of a subsidy—namely, that the subsidy may lead to unanticipated distortions elsewhere in the economy, may require counter-subsidies to offset distortions created, may become burdensome administratively, may inhibit incentives to efficiency, may give unwarranted market protection, and may be difficult to terminate.

The following reading considers some of these elements as related to fertilizer subsidies.

4.12 Evaluating Fertilizer Subsidies

Fertilizer subsidies serve a variety of purposes and take a vast variety of forms. A complete taxonomy is beyond the scope of this paper, but a few of the major characteristics can readily be noted.

Purpose of Subsidies

The major traditional purpose of subsidies in the less developed nations has been to encourage farmers to use fertilizer and thereby expand total production. Since chemical fertilizer is new to many of the farmers in the developing world, most will, so the argument goes, start with a low rate of application on a very limited area. A subsidy is one way of stimulating this process. In terms of the usual S-shaped adoption curve for technology it might be aimed at the innovators and early adopters.

There may be a greater economic need to use a subsidy in the LDC’s than in the DC’s because fertilizer prices are often higher while product prices may be lower. This fact, together with the recent sharp rise in fertilizer prices, may lead to a second and relatively new reason for using subsidies: to help maintain fertilizer use in cases where product prices have not (for a variety of reasons) been raised, or raised correspondingly.

Other reasons for subsidies are more mixed. It has been suggested, on the assumption that fertilizer production exhibits economies of scale, that subsidies may help expand the total domestic market for fertilizers and make the establishment of fertilizer manufacturing and distribution facilities economical. In other cases fertilizer subsidies may be just part of a package to increase food production: in Korea and Bangladesh, for example, fertilizer is part of a program of input subsidies which also includes other farm chemicals, credit, and seeds.

In virtually all of these instances, the unstated assumption is that the subsidy is a temporary measure and that it would eventually be withdrawn. But, as with any subsidy, there is resistance to its reduction from those who immediately benefit, and the subsidy tends to remain much longer than anticipated. At least, it is thought by some, subsidies on fertilizer inputs are easier to phase down than subsidies on product prices.

Policy Alternatives

National fertilizer policies, when they are considered at all, are usually cast in terms of price options or of cost-benefit analysis. A much wider initial focus is needed.

Indeed, one might start with a broader set of questions and follow a deductive process to determine the most appropriate point of program entry. The ordering of the questions might be, for example, as follows:

- What is the major agricultural problem of the LDC?
- If the need is to expand crop production, what are the main constraints?
- If the major constraint is soil fertility, what are the technical options open in the short and long run?
- If chemical fertilizer is the most promising solution to the fertility problem route, what are the major limitations on its use?

The constraints on fertilizer use can take many forms, ranging from physical delivery problems to lack of farmer knowledge about its potential payoff. Of the literally dozens of limitations that might be noted, only a few will be noted here. Initial attention will be focused on the fertilizer cost and product price relationship. This will be followed by a brief review of some non-price policy options.

The two areas are interrelated to some extent. This is because the demand for fertilizer is composed of the relationship between three major components:
(1) fertilizer price, (2) product price, and (3) the nature of the crop's physical response function to fertilizer. The third, though of vital importance, often tends to be overlooked. The response function in turn is related to several other non-price factors, such as the state of technology and the availability of complementary inputs.

The relative importance of these and other components will be influenced by at least three factors. One is the stage of agricultural development with respect to fertilizer. Is the key problem the introduction of fertilizers, or is it one of offsetting current price increases to farmers who are well acquainted with fertilizers? A second factor is the time available; price policies can have virtually immediate reactions, whereas efforts to improve the domestic fertilizer supply and/or the physical response function may take years or decades. A third factor is the stage of economic development of the country: the more advanced and urban nations will have greater financial and technical resources to support such programs than the least developed nations.

**PRICE POLICIES**

As suggested, most of the rather limited analysis of subsidy programs is concerned with the question of the ratio between fertilizer prices and product prices. Since both are inherently variable, the policy questions in many cases are: (1) what balance should be kept as a goal, (2) what approach should be taken to maintain this balance? Obviously in some countries the ratio may be sufficiently favorable that a subsidy is not needed. Where this is not the case, and the ratios have been out of balance, the discussions have been cast in terms of whether it is better to subsidize the fertilizer itself or to exercise some control over the product price. We will briefly note some of the main considerations in each option, and in doing so will cite some of the arguments for a joint approach.

**Fertilizer prices.** The two main ways of reducing relative fertilizer prices in the short run are to (a) lower fertilizer prices through the use of a subsidy, or (b) to raise product prices (or allow them to rise). In the longer run it might be possible to reduce the cost of fertilizer in several ways, such as: importing raw materials and mixing fertilizers, expanded domestic production, and more efficient distribution.

Subsidies on fertilizers, like any public policy, may be expected to have positive and negative social effects of varying strengths beyond their immediate impact on output. An analysis of policy options in Mexico, for instance, has suggested that a 30 percent subsidy on chemical fertilizer might have strongly positive effects on crop production and consumers' surplus, positive effects on exports and producers' income, or a negative effect on the budget.  

Subsidies on fertilizers have been favored in many quarters because, as Krishna puts it, "The benefit of government expenditure can be derived by the peasants only in proportion to their use of improved inputs. Input subsidization also avoids raising food and raw material prices against the growing industrial sector." Subsidy programs is concerned with the question of input price manipulation cannot discriminate between products (some examples of the latter problem have been noted) or income groups of users.  

Fertilizer subsidies have several other limitations. One is the potentially high budget cost to the government, especially if the subsidies are carried beyond the initial adoption period (product subsidies of course share this problem). Another is that if the subsidy is applied to imports, it may discourage the development of a local fertilizer industry.

A more general problem is that subsidies, as Binswanger has noted, may "distort both the choice of commodities and the choice of technique for each commodity." Subsidies, therefore, may encourage uneconomic use of fertilizers; they may be used at the wrong time, in the wrong amounts, on the wrong crops. They may lead to an inefficient choice of cropping patterns.

The subsidy, moreover, may never reach the farmer, or at least the small farmer. Thus, if the subsidy is

---


34. Ibid., p. 527.

35. Schuh felt that this may have been one of the results of Brazil's program of exchange rate subsidies in the 1950s. G. Edward Schuh, The Agricultural Development of Brazil (New York: Praeger, 1970), p. 295.

The relative importance of each would depend on yield due to lower level of fertilizer use, might rise. Still more than pay for itself. Rather, the problem small farmer with limited financial resources. He might, change, and/or the price of food to the consumer, reflecting both the higher input cost and/or lower yield due to lower level of fertilizer use, might rise. The relative importance of each would depend on the nature of the various price elasticities and cross-elasticities of demand.

Factor (1) might be of particular concern to the small farmer with limited financial resources. He might, as suggested earlier, cut back on fertilizer use on the high-yielding varieties (HYV) of wheat and rice. This would not necessarily be because the HYV became unprofitable—studies in Bangladesh, Ghana, and Sri Lanka have suggested that unsubsidized fertilizer would still more than pay for itself. Rather, the problem would more likely be that poor small farmers would have greater difficulty in obtaining credit to buy fertilizer and other needed inputs than wealthier or larger farmers.

With respect to factor (2), the general tendency for fertilizer to flow to the crops with highest returns would be accelerated. In the case of Brazil, for instance, it has been suggested that (a) fertilizer use would fall least on crops like coffee, sugarcane, and cocoa (where it has been used for a long time, where there is a good understanding of fertilization practices, and where product price risk is less due to greater government intervention), and that (b) there would be some shift in cropping patterns (to the above crops and to legume crops like soybeans which generate much of their own nitrogen).

In addition, allowing the price of fertilizer to increase would reduce the incentive for black marketing and smuggling (though the latter might also be influenced by relative exchange rates).

The severity of the effect of fertilizer price increases depends to some extent on the rapidity with which they come about. Part of the current problem is that the prices have increased so much and so quickly that there hasn’t been time to work out rational adjustments. In Sri Lanka, for instance, farm gate prices rose 400 percent in 18 months, leading the country to reimpose a previously discontinued subsidy. On the other hand it was calculated that a 15 percent increase in fertilizer prices, compounded annually, would result in elimination of the existing subsidy in 4 years; a 10 percent increase would eliminate the subsidy in 6 years.

But fertilizer prices are only half of the equation that influences farmer action. The other half is product price.

**Product prices.** Relative product prices, like those of fertilizers, vary sharply around the world. In some areas, product prices are held well below world market levels in order to keep consumer prices, at least in the short run, at artificially low levels. In other cases, prices are supported above world market levels. Clearly an increase in product prices would result in fewer distortions in the former case than it would in the latter. Aside from removing export taxes and the like from domestic production, an LDC government might more actively raise the product price through adjustments in the price support level and/or in official wholesale/retail prices.

The main advantages of raising product prices, compared to input subsidies, are that: (1) they are of major importance and are well recognized by all but self-sufficient (non-market) farmers, (2) a rise, because of the cushion it provides, reduces the danger from downward price fluctuations, (3) a rise rewards the increased use of an array of inputs, and (4) an adjustment can be easily applied to specific or individual crops.

The main operational difficulties of the product price approach center about their lack of linkage to specific inputs. As Krishna and others have put it, if product prices are raised, some peasants “may or may not take to improved cultivation. They may simply spend the extra income on consumption . . .” All farmers will benefit—noninnovators as well as innovators—and there is no assurance that the desired increase in output will be attained.

At the same time, an increase in product price may reward the farmer with higher yields more than the farmer with lower yields. While one effect of this imbalance would be to encourage farmers to increase yields through the use of fertilizer, there may be a host of biological and institutional factors which limit his ability to expand yields. Thus he may benefit less than the farmer who is more fortunately endowed.

The potential for product price adjustments may be limited when an LDC does not have price support programs (or an effective program). Where they exist there may be practical and political limits as to how

---


far product prices could be adjusted to offset increased fertilizer prices. And, on the other hand, it can be politically very difficult to reduce price supports; this is of special importance if the stated product or intent of subsidies was only to stimulate initial use of fertilizers. The government costs, in any case, could be enormous for the effort which is obtained.

If the increased prices are wholly passed on to the consumer, a number of obvious social and political problems would result. Few governments—DC or LDC—care to face the consumer reaction to higher prices or the inflationary pressures they bring about; there is often a strong desire to keep prices down in urban areas. And if prices are raised the nutritional status of lower income groups is apt to be worsened. Still, it may be better to face these problems directly, and to think out ways to ameliorate them (such as subsidized food distribution or direct welfare payments to the very poor) rather than to implement subsidy programs which may be even more difficult or costly in the longer run.

Joint price policy action. It should be quite clear that programs which affect either fertilizer or product prices have their own advantages and disadvantages. One option may rank favorably only when the comparison is made with the other (a fertilizer subsidy may be less expensive than a general subsidy, but still may entail a substantial budget cost). It is doubtful whether there are many locations where either program could by itself be sufficient to efficiently bring about an increase in fertilizer use.

For these and other reasons, the 1967 statement by Millikan and Hapgood that "No blanket preference can in our opinion be given either to general price supports or to input subsidies" still appears valid.39 The relative efficiency of each as a method for improving the price ratio of outputs to inputs depends on many factors which must be examined in the context of each individual country.

Moreover, it is not usually a matter of indifference whether the profitability of a practice is increased by one or the other action. As Krishna has pointed out, there are a number of reasons why the same responses cannot be obtained by manipulating each. In fact, he states, "both are needed as complementary instruments of policy, for different reasons."40

Also there is need to put the matter in a fuller time perspective. In Krishna's words:

It is true that input price subsidization avoids an immediate increase in food and raw material prices, but this will not prevent a long-run step increase in their prices if input subsidization does not succeed in stepping up agricultural output at the same rate as price guarantees would. In other words, input subsidization may seem cheaper than product price support in the short run, but product price support may prove cheaper for the city in the long run.41

For these and other reasons, it may be that where a price policy program is needed, it should be a joint price policy program. And in actual practice, the recent increases in fertilizer prices have commonly led to adjustments in both fertilizer and in product prices. The problem is that they are often not adjusted at the same time; product price increases tend to lag behind those for fertilizer. In Pakistan, for instance, prior to September 1974 "crop prices were not permitted to follow as rapidly and incentives for fertilizer use in major crops declined."42

de Guia has suggested an interesting and relatively sophisticated three-step process which partially integrates the two types of price policy;43

Step I. Introduction (when availability and farmer awareness are of primary importance). Subsidize heavily at the retail point to encourage initial trials with fertilizers. Regulate retail margins.

Step II. Reinforcement (when credit is of primary importance). Subsidize transport costs rather than retail prices. Attack distribution cost structure but permit dealer margins to grow.

Step III. Maintenance (when price is of primary importance). Eliminate subsidies (except possibly where they are used to affect crop-growing patterns). Encourage price competition among dealers to push retail margins downwards.

While it is often recommended that fertilizer subsidies be reduced over time, this has proved difficult. Perhaps a sequence of the above sort, in a period of relative fertilizer price stability, might provide a useful thought framework.

Another alternative might be to have a two-price plan both for fertilizer and product. In the poorer areas, higher subsidies and price supports might be maintained than in the more prosperous areas. The great difficulty, of course, would be to keep the fertilizer from being diverted to other farms and other uses.

41. Ibid., p. 527.
Concluding Remarks

It is difficult to draw many conclusions about subsidies for fertilizers in developing nations. The types of subsidies and range of conditions are too diverse for easy generalization. Still, some comments are possible. These should be prefaced with the acknowledgement that while I had a fairly open mind about fertilizer subsidies when I began this review, I am now considerably more skeptical about them.

The traditional argument for subsidies in LDC has been to encourage fertilizer use. There may well be justification for temporary use of subsidies at the very earliest stages of the adoption process—if it has been determined that a lowered price will provide a significant boost to adoption. This, however, is a significant condition. It is more of a generally accepted nostrum than well-proven fact. Still, it could be true in some cases. And since relatively small numbers of farmers may be involved for only a year or so, the government cost theoretically would not be great.

The problem comes if the government doesn't make it clear in advance that the subsidy is temporary and if it doesn't withdraw it quickly enough. As with any subsidy, political pressures can force its continuation long past its period of peak usefulness.44

The burden carried by the subsidy process has become immeasurably greater with the recent rises in fertilizer prices. These not only make the introduction process more difficult (to the extent that fertilizer price is important), but can add significantly to the cost of production of the adopting farmer. Since the farm demand for fertilizer appears to be more elastic over the long run than in the short run, farmers may let up on their level of fertilizer use over time. The short-run problems, however, may be particularly severe for small farmers with limited credit. Thus there may be some need for financial relief at the farm level if fertilizer, and in turn crop output, is not to drop off.

The question is then whether a fertilizer subsidy is the most efficient and appropriate general vehicle. Subsidies can be very expensive in terms of limited LDC budgets and they do not necessarily insure that the fertilizer reaches its intended use. Increased prices of certain crops could help draw the fertilizer into that use. But to do this entirely by subsidizing the price of the crops could also be expensive. One alternative, where the government has some control over wholesale and/or retail prices, would be to simply allow increased fertilizer costs to be reflected in increased food prices. This would raise the incentive to apply fertilizer (though it would not guarantee its use) and the "taxing" would be done by the higher consumer prices. Since the lowest income groups would suffer in this process, an expanded food subsidy program might be needed to keep nutrition levels from dropping. And additional credit may need to be provided small farmers lest they be unduly disadvantaged.

Some of the funds that would otherwise be spent on fertilizer subsidies could go for an expanded feeding program, but a substantial sum should also be devoted to (1) the development of expanded local supplies, where appropriate, (2) research and education work on increasing the effectiveness of fertilizer use, (3) study of factors which constrain fertilizer use, and (4) study and correction of other factors which constrain the expansion of food output.

Such an alternative program may be unrealistic and open to criticism. It is idealized in that few politicians would carry it out in the face of increased complaints about the retail price. If they did, they might not correspondingly increase the budget for other programs. Also the program might be considered inflationary. But the sad fact is that unless prices at the farm level are allowed to rise, there will be little incentive for farmers to increase output and prices will rise even more.

More general government price subsidies could moderate increased crop prices, but there is always the danger that they might represent an even larger budgetary cost than fertilizer subsidies. This is not to say, however, that there isn't any place where price supports couldn't be useful; a small program might well provide an initial incentive for fertilizer use on certain crops. And if an extensive subsidy program is to be set up, both input and product prices may need to be wrapped up together. Most countries, however, will not be able to afford a highly subsidized program for very long.

Furthermore, there can be a substantial opportunity cost in devoting many resources to subsidy programs. To the extent that policy personnel are engaged in the many complexities of such programs, they may overlook or underfund important other longer-term ways of meeting the fertilizer problem. These ways, in turn, could lead the way for developing a more rationally-based program of fertilizer use on LDC farms in the future.

The problem, as always in the times of crisis, is to escape the present. I see no easy way out. Retail food prices in the LDC are going to have to be allowed to rise to reflect the higher fertilizer cost. To do otherwise by resorting to extensive subsidies is not going to do the trick very long; it is simply too expensive in budget terms. And it does not lay the

44. When this happens, the subsidy could become an agent of income transfer. Subsidies, however, may not be the most efficient or equitable vehicle for carrying out this process.
economic and technical base for continued production increases in the future.

Thus my own view is that subsidies should be used very cautiously and very selectively. There may be a few cases where they can help introduce fertilizer or be of needed assistance to small and/or poor farmers. And there may be instances where adjustments in product prices can help direct their use on certain crops. But on the whole I would tend to allow product prices to rise and, where time permits, turn my thinking and limited funding elsewhere—to the better identification of constraints on fertilizer use, to increasing the efficiency of fertilizer distribution and its use by plants. In short I would try to make the best possible utilization of what little fertilizer I had. This would, however, require much more knowledge about the fertilizer situation at the farm level than presently exists in most developing nations.

Departures from Marginal Cost Pricing

From the preceding readings, it is clear that pricing policy will have implications for both efficiency in the use of scarce resources and investment in providing a greater supply of the scarce resource. We have repeatedly encountered, however, conditions under which efficiency pricing is not feasible or will not fulfill other objectives of cost recovery or income distribution. By way of summary, the final selection in this chapter reconsiders some issues related to departures from marginal cost pricing.

4.13 The Willingness-to-Pay Criterion

This [paper] discusses an old precept of economic analysis that we find frequently troubling economists and others alike when working in developing countries: it is whether, in low-income areas of developing countries, the consumers' willingness to pay for a product is an appropriate basis for deciding on pricing and investment policies. We confine ourselves here to the electric power sector, though the issue is, of course, much wider than this. We do believe, however, that the issue can best be resolved through case-by-case examination of the precept in each of the many different fields to which it is now applied. (We would abandon it, for example, for water supply projects in drought areas of extreme poverty, but would accept it, as will be apparent, for electric power projects.)

Having already spoken of established and new markets, we now distinguish in more detail: established markets, covering (a) nondomestic consumers, and (b) domestic consumers (high and low incomes); and new markets, broken down into cases where (a) incomes and productivity are relatively high, (b) incomes and productivity are low, and (c) incomes and productivity are extremely low. Established markets take the bulk of the investment in developing countries and, relatively speaking, are predominantly prosperous; new markets, in contrast, are predominantly poor. Also, investment to serve low-income markets is only a recent initiative in most developing countries.

Established Markets

The justification for expanding an existing system's supply to an area already served rests largely upon the gains that can be provided to individual consumers through increased supplies. There may also be some community benefits from public lighting. But most of the benefits are gained by individual consumers, domestic and nondomestic. Consider nondomestic consumers first.

Increased industrial use of electricity is a necessary condition for industrial growth. For lighting, small motive power units, and some chemical processes on the one hand, the substitutes for electricity are so inferior that any check to availability would retard the expansion of industrial output. For heating and large motive power units, on the other hand, firms can often use coal, gas, or oil as alternative sources of energy. The choice is one of cost-effectiveness; if firms are to make choices that are sensible from a national economic viewpoint, the prices they face should reflect the national economic costs of supply. This requires, among other things, that industrial electricity tariffs have a structure, for example, as between peak and off-peak power, that reflects the structure of supply costs. If, in addition, the general level of electricity tariffs relative to the price levels

of other forms of energy reflects their relative costs as seen from the national economic viewpoint, the cost-saving incentives presented to firms will lead them in the direction of making sensible choices. Thus it makes sense to allow the rate of expansion of industrial electricity consumption to be determined by firms' own choices.

Consumers equip themselves to use the public supply at considerable expense. Expenditures on electric motors, for example, may exceed the costs of purchasing electricity to run them. Once the market is established and is adapted to public services, maintaining the continuity and the quality of service becomes important. A deterioration in the continuity of electricity supply due, for example, to large backlogs in the investment program, disrupts business activity on a large scale. Where this has happened in some countries, firms have been driven to installing autogenerators to maintain service, even though the costs of doing this are greater than the costs to the country of expanding public supply. Continuity and quality of supply to industry and commerce in expanding markets need to be maintained, provided that prices reflect costs.

It is with respect to the domestic consumption of electricity that difficulties arise with the idea that, provided tariffs reflect costs, the pace of expansion should be determined by the growth of demand. The use of more electricity by higher-income groups in developing countries may well be considered less important than, say, housing for the poor or new rural roads. But, in accepting this, it is necessary to recognize that simply refusing to meet the demands of the higher-income groups would create problems. Some mechanism for transferring resources must be introduced. One such mechanism is to provide the electricity but to tax it. This would simultaneously dampen demand and provide revenue for expenditure on other development projects.

Otherwise, simply depriving the higher-income groups of service leads to all kinds of difficulties and inequities. In some electric systems, the only way to ration supply is to shed it in a manner that may also entail the shedding of loads to industry and to low-income consumers (because they are connected to the same network). Thus, physical deprivation of services is usually inefficient and often inequitable. Finally, a useful basis for raising the public revenue is lost.

It is better, therefore, to meet the growing demands of higher-income consumers than to ignore them. Taxes can be introduced if appropriate. The choice of taxes is a broad issue, of course, and cannot be settled solely within the context of public utility projects. There may be a better case for taxing other goods. But once the tax policy is decided upon, there is no other sensible alternative open to the utility than that of meeting the demands of all consumers at a satisfactory level of quality and continuity of service.

There are, of course, some cases where a project that raises the capacity of an existing system to meet demands in existing markets can be postponed temporarily. These instances occur either when the adverse effect upon the quality of service is marginal or when prices that are too low can be raised sharply, with the prompt effect of checking the growth in demand. But these cases are uncommon: the quality of service is seldom too high, and the short-term effect of prices upon demand is usually very limited. Adjustments to higher prices often require new plans and new equipment and appliances and so take place gradually.

In general, therefore, continued system expansion is necessary. But this most emphatically does not signify that public utilities should be allowed to display an unlimited demand for capital. The proposition that system expansion should not be stopped suddenly in the short run leaves open the question of its rate in the long run. It is here that the real options lie.

New Markets

A significant portion of the increasing demand for electricity in developing countries stems from newly constructed areas in cities, housing relatively high-income families or being centers of expanding industrial, commercial, and governmental activity. Much of what has been said above applies to projects to serve such new areas. The projects are worthwhile if consumers are prepared to pay prices that reflect the costs of the projects and their outputs, plus any taxes that are considered appropriate. Since many similar consumers are supplied currently, there is generally ample precedent for the demand forecast. A comparison between forecast revenues (including taxes) and the opportunity costs of the project and its output will testify to the project's desirability (if tariffs reflect costs plus any appropriate taxes). If the revenues are below opportunity costs, this is almost certainly because the tariffs are too low.

In all new markets—and this is a general point—an alternative to simply accepting a project may be to offer service at a lower quality in order to supply lower-cost service to more customers at lower prices. But as noted above, it is important that the utility should not reverse decisions suddenly and significantly reduce the quality of supply once the market has become established and has adapted to the quality of supply first agreed upon.

Major difficulties arise in the extension of service to low-income areas in villages and cities where do-
In addition to these benefits, there may be a benefit to the community from public lighting, as mentioned earlier.

Because of the importance of these benefits, a satisfactory return on the investment may often be expected. Electricity is a factor input in nondomestic consumption in industry and agriculture, and this indicates the kind of returns that can be expected. When service is desired by nondomestic consumers, it is because it contributes to profits and, apart from the smaller of these consumers, supply costs can be met. Moreover, it is desirable that revenues do meet costs. For example, in irrigation (which shows good profits) a particular problem is to encourage efficiency in the choice between diesel and electric pumping; this can be done if the prices of electric pumps and of electricity reflect the costs of supply, on the condition, of course, that the prices of diesel pumps and energy also reflect the costs of providing them.

Furthermore, since incomes are not uniformly low, a number of households can often afford and are indeed prepared to meet the costs of service. Where electricity serves productive purposes, and where enough domestic consumers are in this "less poor" category, prices can be set to reflect costs and a satisfactory return to the investment can be expected. If the returns are low, this may be because the tariffs are too low or because the quality is too high (and costly). In this case there are two alternatives to simply accepting the project as it is: (a) raise tariffs, or (b) lower the quality (and costs) of the service for some consumers. Alternatively, if it is desirable to assist low-income business and domestic consumers in the area, a degree of subsidy may well be acceptable. Such subsidy could take the form of, for example, a low first block in the kilowatt-hour tariffs, or reduced connection costs for low capacity supplies.

The central precept suggested for the economic evaluation of electrification projects is that the desirability of these projects in the circumstances so far described is revealed in the consumer demand at prices that are a compromise between fairness and efficiency. Even in low income areas—say, down to about $50 per capita—where development is just beginning, it seems that this precept holds quite well. In fact, a major obstacle to a successful investment seems to be that the demand for the services generally is underestimated and that the factors that relate to successful investment are not well understood. These include the management of prices, subsidies, and credit policies; the institutional framework; the promotion of and institutional interest in the service; proper monitoring and evaluation of project performance; the indirect influences on the project of public investments and support for local agriculture and
industry; and the tradeoffs between quality of service and the number of people who can be supplied.

In areas of poverty and extremely low productivity, however, the precept of relying on people's demand for the service and their willingness to pay for it at "fair and efficient" prices breaks down, for the simple reason that these people lack the incomes to purchase service. What can be done in such circumstances? One answer, though it has obvious defects, is for the government to subsidize the service very heavily. In the case of electricity, this does not seem very useful; for the service to be of any use, electricity requires substantial complementary investments in electrical machines and apparatus. Aid to extremely poor areas requires a broad program of investments and economic and institutional reforms and initiatives. Electrification projects can form no more than a part of such programs.

With the exception of such extremely poor areas, and subject to the possibilities of taxes and subsidies mentioned above, our conclusion is that consumer willingness to pay for electricity is an adequate monetary measure of its economic benefits, and a good basis for pricing and investment decisions.
Economists have increasingly urged the governments of developing countries to "get prices right." It is to be hoped that the concepts and principles that we have identified in the foregoing chapters might provide a better understanding of how this can be achieved. We have seen repeatedly that appropriate prices require appropriate policy measures to improve the efficiency of markets—whether they be for labor, foreign exchange, agriculture, or public products. Market imperfections and price distortions can exist in any type of market, and public pricing policies may be needed in many markets. The concepts and principles that we have examined should therefore have wide relevance throughout the economy. This final chapter offers some general conclusions that follow from the concepts and principles that we have considered.

General Conclusions

The broadest conclusion points to a need to extend market institutions in a developing economy and to correct some policies and introduce others in order to allow the price system to operate more effectively. Price changes are most important in determining resource allocation and resource mobilization. Reference to prices is indispensable in reaching rational decisions on what to produce and in selecting the means by which output is to be increased. Beyond static resource allocation, the price system also has dynamic consequences for the mobilization of resources to increase output. Appropriate pricing can mobilize additional savings, increase the demand for labor, and support structural changes involving agricultural development and export promotion.

In emphasizing some fundamental policy implications of the price system, we have indicated that pricing is not an activity of the accountant but of the economist. The economist's concept of cost is fundamental: costs must be defined in relation to their opportunity cost, which is the forgone benefit of not using the resources in the best alternative activity.

Once we recognize the economist's notion of cost, we can readily understand the various functions of a price system. A price system is characterized by specialization and division of labor, exchange transactions on product markets and factor markets, the exercise of choice by economic agents among alternatives, and the adjustment to changes in prices that result in changes in inputs, outputs, demand, and income distribution. The more developed a country's price system, the more effectively it will exercise the functions of providing information, rationing, allocation and mobilization of resources, and distribution of income.

For a developing economy, it is also important to recognize how the market-price system can be used for the exercise of public policy.
Many economists have come to believe that in order to overcome past deficiencies in planning and undertake policies that conform more closely to their present needs and capabilities, the majority of LDCs should retrench to a lighter type of planning. This would rely more on decentralized decisions operating through the market mechanism, and greater attention would be given to devising policies that might make private action more effective.

The advocacy of greater use of the market mechanism should not be interpreted as a call for a diminished role for the government, but rather a different role. Governmental policies are needed to strengthen the market system, and a stronger market system is in turn needed to allow public policy to operate more effectively through the market. Thus, in a sense, more planning is actually required to overcome the results of inadequate planning in the past: improved planning is now necessary to remove the distortions caused by arbitrary direct administrative controls that have produced a disequilibrium system and a set of trade, fiscal, financial, industrial, and wage policies that are often contradictory and self-defeating.

The case for a high degree of planning, with a large amount of public investment and deliberate industrialization constituting the core of the plan, may remain strong for countries that suffer seriously from the pressure of population on the land but have the potential for large domestic markets (such as India or China). For other countries, however, in which there are underutilized natural resources but only small internal markets, it can be persuasively argued that the basic problem is one of creating a favorable social and economic environment which will lead to expansion of private activity, more effective use of the underutilized resources, and capitalization on the existing opportunities for international trade.

If the private sector is to be enlarged, how is this now to be accomplished? Instead of allowing any economic resources to be left unemployed, governmental policies must try to mobilize through positive economic incentives and inducements the latent skills and capital in the private sector. Fundamentally, the stimulation of multiple centers of initiative depends on the establishment of markets and encouragement of market institutions. Economic and social overhead capital can help to establish the physical conditions for a market to exist and can support the interdependence of markets. The government also has a crucial role in building institutions such as a banking system, a money and capital market, agricultural cooperatives, labor organizations, rural credit institutions, and training institutes. It must also be recognized that many policy measures that can affect individual action by altering the economic environment are not of the usual monetary or fiscal type of policy, but rather of a kind that involve the legal and institutional framework, such as land tenure legislation, commercial law, and property rights.

Once market imperfections are reduced and the structure of markets improved, the market mechanism can itself be used as an instrument of development—promoting governmental policies as well as more effective private activity. Instead of relying on comprehensive and detailed administrative controls, the government can alter prices to execute policy and can provide price and income stimuli for an expansion in private output, an increase in exports, and a widening of domestic markets. These price changes may extend to foreign exchange rates, interest rates, tariffs, taxes, and subsidies. Subsidy and tax schemes can be especially relevant in inducing firms to value inputs according to their social opportunity costs, to exploit external economies, or to introduce new techniques of production.
In Chapter 3 we outlined the efficiency conditions that would result in perfectly competitive markets—not that such markets actually exist, but as an ideal reference standard against which we can compare the actual market conditions and then consider remedial policies. We established that an efficient economy would achieve technical efficiency ("be on the production possibility frontier") and economic efficiency ("Pareto optimality" for a given income distribution). Given perfectly competitive conditions, and the freedom to make voluntary exchanges in the market, it follows that a voluntary exchange between two persons that affects no other person leads the economy closer to Pareto optimality, for it makes one of the individuals better off, in terms of individual preferences, without making anyone worse off. If all markets were in perfectly competitive equilibrium, then producers would be equating price to marginal cost and to minimum average total cost, realizing only normal profit, and consumers would be "maximizing their satisfaction." From the standpoint of resource allocation throughout the economy, the rate at which one product can be transformed into another in production (the marginal rate of transformation on the production possibility frontier) is equal to the price ratio and to the rate at which the commodities can be substituted in consumption (the marginal rate of substitution). When this equality holds, no transfer of resources is possible between any pair of commodities that will increase the supply of one of them more than just enough to compensate consumers for the reduction in the supply of the other.

Although our exposition has necessarily taken us into some detailed technicalities of the price system, we must not lose sight of the general social significance of the price system. As one economist expresses it,

Our concern is with social decisions. Any decision is a choice among alternatives. Social decisions can be distinguished from technical decisions on the basis of the criteria used to evaluate the alternatives. A technical decision is one whose alternatives can be rated objectively according to the amount they contribute to attaining some single, well-defined goal. For example, if a river is to be dammed as cheaply as possible, the selection of the site at which construction cost is lowest is a technical decision. On the other hand, a social decision is one whose alternatives affect the attainment of a number of different goals, so that the choice among alternatives depends on the relative importance of those goals. There is no purely objective way to rate the alternatives in a social decision. For example, if the cheapest site for the dam is also the one most destructive of scenery, and if scenery has social value, then a social decision is involved in the site selection. A single consumer's division of his budget among commodities is a social decision, according to this definition. Every decision that affects a number of people is a social decision, since the well-being of every individual is a distinguishable goal.

The essence of a social decision is that it impinges on a number of goals which are likely to conflict. If several people participate in the decision, they need some mode for expressing the relative importance they attach to the different goals. Prices frequently fill this need. For this reason, prices have been called "coefficients of social choice." When prices are applicable, we have an economic decision. Economic decisions are easier than other types of social decisions simply because they benefit from this peculiarly efficient method for expressing and communicating individual preferences.

The circumstances in which prices are helpful in arriving at social decisions are complicated. But one aspect is clear at the outset: a price does not pertain to a social goal, but to an instrument for attaining it, called a commodity. This brings us back to the concept of opportunity cost: the price of a commodity measures its potential contribution to the attainment of the different goals to which it is applicable. A commodity should not be used for any purpose unless its contribution to
the purpose is at least as important as its potential contribution to other social objectives, as indicated by its price. In this way, reference to prices can allocate commodities to the uses in which they make the greatest social contribution. This, in essence, is the social function of our system of prices and markets.1

The foregoing quotation, together with the general theme of this book, emphasizes the importance of prices in the development process. But some students of development have cautioned that this emphasis can be overdone. Thus, in comparing three strategies to achieve simultaneously the policy objectives of elimination of poverty, reduced inequality in income distribution, and an increase in the growth rate of the income of the poor, Selection 5.1 poses alternatives to the traditional emphasis on pricing policy.

5.1 The Price Mechanists

One may distinguish, very broadly, between three schools of thought, each advocating a different strategy to eradicate poverty and reduce inequality. They may be called, for want of better names, the Price Mechanists, the Radicals, and the Technologists. The Price Mechanists argue that low production, low productivity, inequality and unemployment can be eliminated by setting the correct prices, which serve both as signals and as incentives.

There is a powerful and vocal group of development economists who argue that many, if not all, of the disappointments with development efforts are due to faulty price policies. Governments have set the wrong price for capital (too low and often rationed, encouraging excessive scale and underutilization of capital, discriminating in favour of large firms and encouraging take-overs of local by foreign firms), for labour (too high, contributing to unemployment and underutilized capacity and discouraging exports), for the foreign exchange rate (overvalued, discouraging labour-intensive exports and encouraging high-cost import substitution), for the products and services of public enterprises (too low, subsidizing the private modern sector). If only governments were to set the right prices, economic growth, as well as jobs and justice, would triumph. Indeed, this group argues, many of the evils attributed to foreign investment, the multinational enterprise, the wrong technology, inappropriate products, the dominance of the developed country and the dependence of the under-developed country, the terms of trade, international inequality, etc., are really due to "distortions," to faulty pricing policies, which convey the wrong signals and provide the wrong incentives.

Most people would agree that "getting prices right" is not enough. Some would say it would go a long way towards combining more growth (and more efficient growth) with greater equality, others would say that the contribution would be only marginal, but all would agree that other things would have to be done as well, if only additional marketing efforts for the extra exports generated by the "right" price policies, or better facilities to improve the capital or labour markets. (But some would argue that the "right" prices themselves would provide sufficient incentives for the creation or improvement of these institutions.)

But to say "getting prices right is not enough" is open to two diametrically opposite interpretations. It might mean either that, by itself, it would make a contribution to the eradication of poverty and to greater equality, though this contribution would be greater if other things were done as well. But getting prices right is better than nothing. Alternatively, it might mean that while correct pricing policies combined with structural reforms, and in particular the redistribution of assets, would contribute to growth and equality, by themselves they might make matters worse or simply alter the manifestation of inequality.

No doubt, the "wrong" price policies can impede development, reduce employment, strengthen monopoly and aggravate inequality. But it does not follow that the "right" policies necessarily do the opposite. They might simply lead to different forms of the same evils. Let us assume that land and real capital equipment are scarce, while unskilled labour is plentiful. The supply of labour is growing faster than that of land and capital. We allow competition to prevail and factor rewards to be determined by marginal productivity. And we allow for a fair degree of substitution between labour and capital. Then rents per acre and real returns on capital will be high and rising (reflecting the growing relative scarcity of land and capital), while the wage rate will be low and falling (reflecting the growing abundance of labour).

Producers will pay much to get hold of scarce resources of land and capital equipment and will offer little for the abundant supply of workers seeking jobs. Low wage costs will tend to expand employment and output, the extent depending on the elasticity of substitution. Processes, sectors, and products that are labour-intensive will be encouraged and those requiring land and capital discouraged. Foreign capital will be attracted from higher-wage, lower-profit countries. High profits and low wages will tend to encourage domestic savings and hence increase the supply of capital inside the country. Moreover, there will be incentives to invent new methods and products that use labour and save land and capital. All this is fine and as it should be.

But these desirable incentives depend upon wages being low (and, if labour grows faster than land and capital, falling), while rents, interest and profits are high (and rising). If in such a society the distribution of land and capital were to be very equal (peasant proprietors or socialized ownership of means of production, though the latter would raise questions as to how the state determines wages and the disposition of the surplus and also about the unequal distribution of power), the functional differences would not matter because personal or household equality would still prevail. What a family loses on labour income, it gains on property income. But if property (land) distribution is unequal, if the ownership of assets, including access to educational opportunities, is highly concentrated, inequality might increase, even if the share of wages went up as a result of a fairly high elasticity of substitution. The difference between a high-wage, low-interest and a low-wage, high-interest policy is that in the first case the evil takes the form of unemployment, in the second of inequality between wage earners and property owners. "Getting prices right" may therefore transform inequality within the working class (between those with and without jobs) into inequality between workers and owners of assets.

To conclude: correct pricing is certainly not enough where ownership of assets is concentrated and, by itself, may make matters worse. This does not mean that correct pricing, combined with other policies, has not an important part to play.

This brings us to the second school of thought: the Radicals. This school believes that what matters is to redistribute assets, power, and access to income-earning opportunities. Only through such "structural" and institutional reforms, whether peacefully and gradually or through revolution and quickly, can growth and equality be achieved.

Both the Price Mechanists and the Radicals are (often implicitly) optimistic about technology. They believe either that the technologies appropriate for the eradication of poverty and for the promotion of greater equality already exist, or that the "right" prices or the redistribution of assets will automatically provide the incentives to invent them. The third school, the Technologists, are not so optimistic.

They approach the solution of the problem of poverty, unemployment, inequality and low productivity like that of putting a man on the moon, or, nearer home, discovering new high-yielding varieties of wheat, maize, and rice. Industrial technology, public health, low-cost housing, birth control, nutrition, crops for small farmers, and urbanization require the concentration of brain power and research resources. The Technologists are right, insofar as neither prices and incentives nor "structural changes" can solve a problem where the appropriate technical solution just does not exist: where it has to be invented or discovered and where the incentives are too weak or too slow-working to produce "automatically" the right solution.

The conclusion of the discussion is by now plain. We have an instance of the Theory of the Second Best, according to which $\alpha + \beta + \gamma$ yield the desired results, but $\alpha$ or $\beta$ or $\gamma$ by themselves, far from being "better than nothing," may move society away from the desired goal. Only a three-pronged attack, combining signals and incentives, institutional reforms directed at the redistribution of assets (including access to education), and technical and institutional innovation promise results. The precise combination of price policies, asset redistribution, and technological research will depend on a number of factors that will vary between countries: on the readiness of vested interests to yield, on the elasticity of substitution between factors, on the nature of the interdependence between sectors, on the degree of concentration of ownership, on the productivity of assets when redistributed, on the fiscal system, etc. But it is only on the three legs of this tripod that efficient redistribution can rest.

It may be objected that to demand a simultaneous attack on all three fronts is to ask for the impossible. This, it might be argued, is suggested by the frequent failure to meet any one of the three requirements. However, the relations between the three are such that failure to attempt change on one front only may prevent the change itself, or the desirable effects of the change. Hence the failure of one-prong attacks may actually be due to confining attention to one aspect.

However, much may depend on the order in which the policies are pursued. Taking one sequence may prevent the achievement of any or all of the targets. For example, price changes preceding income redistribution may establish patterns of production and consumption inimical to redistribution. Yet if redistribution precedes policy changes on technology and prices, it may establish pressures likely to bring them
about. Irma Adelman concludes that the sequence of reform is critical to the relative success of some countries which have combined improvements in the incomes of the poor with accelerated growth.

Irma Adelman has argued that an "examination of the development process of those non-Communist countries which have recently successfully combined improvements in the incomes of the poor with accelerated growth (Israel, Japan, South Korea, Singapore and Taiwan) shows that they all followed a similar dynamic sequence of strategies ..." Her strategies do not correspond precisely to ours, but the sequence is first, radical asset redistribution (sometimes accompanied by negative growth rates); secondly, massive accumulation of human capital and skill creation far in excess of current demand; and, thirdly, economic policies directed at rapid, labour-intensive economic growth, with the development of appropriate technologies for large countries and foreign trade for small countries. Historically, the "grow now, redistribute and educate later" strategy has been followed by some capitalist economies, but the "redistribute and educate now and grow later" strategy is the one followed by the economies studied by Irma Adelman. It is more consistent with current demographic trends and time scales.

In considering the correct sequence one must pay attention to the pressures which reform on one front will have in inducing or preventing reform on the other two, and the likely impact of the one reform taken by itself. From this point of view, it seems that radical redistribution should come first, then the other steps may follow. This reverses the sequence in the now developed countries which was "grow first, redistribute later." This took many decades to make the poor better off and bring about greater equality; and it contrasts with the "redistribute marginally and grow" policy, or "redistribution with growth," which is likely to fail in one or both aims.

Policy Instruments

From the general perspective of development programming, Selection 5.2 indicates a variety of public policy instruments that operate through prices. The range extends from general monetary policy that affects the price of money (the interest rate) to specific policies that affect the prices of commodities (through taxes or subsidies) or the price of a factor of production such as labor (through a wage subsidy). The superiority of operating through price variables rather than through quantity variables (direct, quantitative controls) is also emphasized.

5.2 Choice of Policy Instruments

Policy instruments may be classified in various ways: by the sectors of the economy on which they operate, by their use of prices or quantities as variables to be manipulated, by the extent to which they can be effectively controlled by the government, by the effect that they have on private incentives and freedom of choice, etc. In table 5-1 representative instruments are classified according to the extent of their application (general versus specific) and their mode of operation (through prices or quantities). The general instruments act on broad aspects of the economy—the money supply, the government budget, investment, consumption—and are widely used in developed and under-developed countries alike. The specific instruments are applied differentially to individual sectors of the economy, as illustrated by subsidies, tariffs, or government investment.

To achieve a given effect on production, or use of any commodity, there is a choice between controlling a price and controlling a quantity. In this respect, tariffs are an alternative to quotas, differential interest rates are an alternative to capital rationing, and subsidies to private producers are an alternative to production by the government. These measures differ in their effects on prices and consumer choices, in administrative convenience, in the predictability of their results, and in other respects. A choice between quantity and price variables as instruments must therefore be made by balancing the advantages and disadvantages in each case.

Some of the main issues of economic policy are concerned with the choice between general and specific instruments and between using prices and quantities as control variables. There is a strong case to be made for using general instruments rather than

Table 5.1. Classification of Policy Instruments

<table>
<thead>
<tr>
<th>Area of policy</th>
<th>Instrument</th>
<th>Variables affected</th>
<th>Quantity variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monetary</td>
<td>Interest rate</td>
<td>Level of investment, Cost of production</td>
<td>General open market operations, Money supply, Prices</td>
</tr>
<tr>
<td>Fiscal</td>
<td>Personal income tax</td>
<td>Consumption and saving</td>
<td>Government expenditure, National income, Price level</td>
</tr>
<tr>
<td></td>
<td>Corporate income tax</td>
<td>Profits, Investment</td>
<td></td>
</tr>
<tr>
<td>Foreign trade</td>
<td>Exchange rate</td>
<td>Cost of imports, Price of exports, Balance of payments</td>
<td>Exchange auctions, Exchange rates</td>
</tr>
<tr>
<td></td>
<td>General tariff level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign investment</td>
<td>Taxes on foreign profits</td>
<td>Level of foreign investment</td>
<td>Foreign loans and grants, Investment resources, Exchange supply</td>
</tr>
<tr>
<td>Consumption</td>
<td>General sales tax</td>
<td>Consumption</td>
<td>Social insurance, relief, other transfers, Consumption income distribution</td>
</tr>
<tr>
<td>Labour</td>
<td>Wage rates</td>
<td>Labour cost, Profits and investment, Labour income</td>
<td>Emigration and immigration, Labour supply</td>
</tr>
<tr>
<td>Production</td>
<td>Taxes and subsidies, Price control</td>
<td>Profits and production</td>
<td>Government production, Level of production, Government research and technical assistance, Cost of production</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Investment</td>
<td></td>
</tr>
<tr>
<td>Investment</td>
<td>Interest rates, Tax exemptions</td>
<td>Profits, Investment by sector</td>
<td>Government investment capital rationing, Level of investment, Restrictions on entry, Prices and profits, Level of investment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumption</td>
<td>Specific sales taxes</td>
<td>Consumption by commodity</td>
<td>Government services (health, education), Consumption income distribution</td>
</tr>
<tr>
<td>Trade</td>
<td>Export subsidies, Tariffs</td>
<td>Price to consumer, Profits on domestic production</td>
<td>Import quotas and prohibitions, Level of imports, Domestic prices</td>
</tr>
<tr>
<td>Labour</td>
<td>Wage subsidy</td>
<td>Labour cost and use, Profits and investment</td>
<td>Labour training, Supply of skilled labour</td>
</tr>
<tr>
<td>Natural resources</td>
<td>Taxes and subsidies</td>
<td>Cost of production, Rate of exploitation</td>
<td>Surveys, auxiliary investment, etc., Rate of development</td>
</tr>
</tbody>
</table>

a. All taxes affect government revenue and saving in addition to the variables cited.

specific ones. The rates of interest, taxation, and exchange are the orthodox means of exerting government influence in a laissez-faire economy. Their immediate objectives are stability in prices and the balance of payments and the prevention of unemployment. Growth is left to free market forces. The manipulation of interest rates and exchange rates allows market forces in each sector to determine where expansion or contraction of production and consumption will take place. These instruments therefore interfere less with the choices of producers and consumers than do measures which discriminate by
sector. They also require a less detailed analysis for their use and do not substitute government judgment of what is desirable for the action of market forces.

The need for specific instruments to supplement general measures derives from the deficiencies in the price mechanism which apply primarily to specific sectors of the economy. When these factors prevent the achievement of a satisfactory rate of growth, the problem is to devise policy measures which will improve on the working of the competitive economy without losing the advantages of private initiative and the automatic adjustment of the price system.

In designing policies for specific sectors, there is an argument for using price rather than quantity instruments which is based on reasoning similar to the case for general over specific instruments. Taxes and subsidies distort the choices open to producers and users of a commodity less than do allocation systems or other quantitative restrictions and hence are conducive to greater flexibility and overall economic efficiency. Furthermore, the administrative requirements for price intervention of this type are generally less than for quantitative controls.

Despite the general case in favour of using the price system, there are several situations in which quantitative measures may be needed:

(i) When it is necessary to limit consumption of an essential commodity in short supply (e.g., imported goods), the tax needed to bring about a given reduction in use might result in such high prices that the burden of the reduction would fall on lower income groups. In this case, price controls and rationing may be preferable on welfare grounds.

(ii) Where a minimum increase in production is essential to production in other sectors—as in the case of power, transport and various auxiliary facilities—the price needed to ensure adequate private investment may be too high or the response of private investors too uncertain. In this case, quantitative measures, such as government investment, may be more efficient because the cost to the society is less or the outcome more predictable.

(iii) In general, where controls are needed for only a short period, as in the case of temporary shortages, it may be desirable to allocate supplies to more essential uses rather than upset the general price structure and distort investment decisions by allowing prices to rise. Quantitative measures are also likely to have more predictable effects in this case.

In these examples, it is the dynamic elements in the situation and the deviation from a desirable income distribution which provide the principal arguments for using quantitative measures of control.

Specific measures for investment allocation: Although the specific measures listed in table 5-1 affect both current production and the allocation of investment resources, it is the latter aspect that is crucial for the future course of development. The various instruments affect investment decisions through the availability and cost of primary inputs (labour, natural resources, imported commodities); through the supply of inputs from other sectors (raw materials, overhead facilities); through the demand for output (sales taxes, export subsidies); through profits (taxes, subsidies); and through measures directly related to the process of investment (interest rates, capital rationing, restrictions on entry, direct government investment). There is therefore a considerable variety of choice between quantity and price instruments and among measures more or less directly related to a particular investment.

The a priori arguments concerning some of the principal measures for influencing investment decisions run somewhat as follows:

(i) Measures of protection. Protective devices are perhaps the most common instruments for influencing the pattern of investment. For this purpose, tariffs are generally preferable to quantitative restrictions—quotas, prohibitions, exchange controls, etc.—for reasons already indicated. Quantitative restrictions prevent competition with domestic producers regardless of price, raise prices to users and limit demand, and require an elaborate administrative mechanism and detailed economic analysis to be effective. Quotas also involve a loss of revenue to the government, as compared with the use of tariffs, unless the profits of importers can be recovered through taxes.

The cases where quantitative measures may nevertheless be needed derive from the principles given in the previous section. In cases of extreme shortage of foreign exchange, tariffs (or devaluation) may be too uncertain in their results and quotas or exchange restrictions may be adopted as emergency measures.

The effect of quantitative restrictions on investment in domestic substitutes for imports or in sectors using imported commodities is generally less certain than that of tariffs. Allocations are subject to variation according to the amount of exchange available, and the profitability of domestic production is harder to determine than in the case of a tariff.

As instruments for inducing investment in new types of production, subsidies may be preferable to either quantitative restrictions or tariffs because the price is not raised above the level of world prices. Total demand is therefore greater and using sectors are not penalized in export markets. The cost of this technique in government expenditure must be weighed against its benefits, however.

Protection from foreign competitors is only one factor in the expansion of domestic production. Also required are entrepreneurs, capital, skilled labour, raw materials, etc. When some of these are lacking, the restriction only serves to reduce imports and raise prices to consumers. Trade restrictions are therefore a rather uncertain method of directing investment unless combined with other measures affecting factor
supply, and they frequently have undesirable secondary effects.

(ii) Government investment versus incentives to private investment. Although the arguments concerning trade restrictions are based mainly on economic considerations, the choice between government investment and incentives to private investors involves social and political factors to a large extent. In countries that do not have strong ideological preferences for either private or government enterprise, the usual approach is to rely on private investment except in cases where it cannot be expected to work in the public interest (e.g., monopoly) or in which its performance has been demonstrably deficient. Since the reaction of investors to various incentives (tax reduction, guaranteed markets, low interest rates, etc.) is subject to considerable uncertainty, such incentives are more likely to be adequate when a general objective is to be achieved—e.g., import substitution, increase of industrial employment—than when increases in output in specific sectors are required. Because of this uncertainty, the extent to which reliance on private investment is desirable can be determined only by an actual trial of specific measures.

Another alternative for securing investment in given sectors when tax incentives are thought to be inadequate or too costly to the treasury is the intervention of a government agency as entrepreneur but not as a long-term producer. This may be done through development corporations, which sell their investments to private enterprises as they become profitable, or through mixed corporations, in which the role of the government declines as the enterprise becomes established.

The assumption underlying all these measures is that it is bad for the government to continue permanently as a producer in most fields. There is a widespread view (shared by the present writer) that the lack of incentives to efficiency in government operations makes private operation preferable even where conditions are not favourable to the initial undertaking of the investment by private enterprise. In the absence of more objective evaluations of the experience with government and private enterprise in various countries, it is impossible to support this conclusion empirically, and it is by no means universally held among democratic governments. In countries such as India, for example, an attempt is made to ascertain the relative merits of public and private investment in specific fields rather than starting from this premise. Even in these cases, however, the sectors that are chosen for government investment are limited in number and characterized by specific structural features (economies of large scale production, importance of the product, tendency to monopoly, etc.).

The possibility of attracting foreign investment adds a further element to the problem. To the argument against government investment must be added the loss of additional investment resources, while the argument against private foreign investment must include the removal of profits from the economy and the future burden on the balance of payments. A purely economic evaluation would probably weigh the value of the additional investment resources and managerial talents more heavily than the cost of obtaining them (particularly where there are unemployed labour and natural resources because of lack of these factors), but the decision is infrequently made on purely economic grounds.

Quantitative analysis and choice of instruments: The preceding discussion has been entirely in qualitative terms, which at best leads to the establishment of certain cases to which particular policies apply. The identification of an actual situation with the relevant case often depends on the results of quantitative analysis. Such factors as the extent of the excess demand for imports, the future amount of unemployed labour, the magnitude of the shift in resources needed in particular sectors, and the importance to the rest of the economy of a given investment, can only be determined from such an analysis. The initial study of development possibilities should be designed to permit a choice of policy instruments in different fields. Once this has been done, the long-term programme can be formulated in more specific terms which take account of the instruments chosen.

The importance of a quantitative analysis for the choice of policy instruments will be determined in part by the presence or absence of the following factors:

(i) Economies of scale in production;
(ii) The possibility of imports and exports;
(iii) The use of the product in other sectors of production;
(iv) The predictability of demand.

In the production of consumer goods, the main objective of the development programme is likely to be a certain degree of substitution of domestic production for imports, but the choice of sector can be left to market forces. Quantitative analysis may be needed to determine the amount of employment and exchange saving which should be aimed at in the consumer goods industries, but not to determine the choice of sector.2

At the other extreme, the amount and distribution of investment in overhead facilities must be deter-

2. This statement is not true where economies of scale are important as in the case of automobile production, because then the profitability of investment depends on an estimate of the quantity that would be demanded at the expected level of income.
mired entirely from a quantitative analysis of future production because the alternative of imports is not available and output is needed to permit investment and production in other sectors. In some cases the choice between public and private investment will also depend on the amount of output required.

Choices among policy measures in the intermediate goods sectors are more affected by the outcome of the quantitative analysis than are those in consumer goods because demands derive from the planned outputs of the using sectors. Economies of scale are also more prevalent, and there is thus more interdependence among investment plans in earlier and later stages. While imports provide alternative sources of supply for many intermediate goods, some investments will not be undertaken unless there is a domestic supply of materials available. To ensure the carrying out of several interconnected projects, government intervention in some form is likely to be necessary because the risk to private investors would be too great. Investments centering on steel production—ore, transport, power, iron and steel, fabricating—provide a good example. Once the initial investments have been made, however, most of them will prove suitable for private ownership and operation.

The advantage to the economy—in terms of the social productivity of the total investment—of interrelated projects of this type cannot be accurately determined from a partial analysis of each investment taken separately because the profitability of one may understate its contribution to the total. This dynamic type of external economy (as opposed to the technological external economies of static analysis) can only be taken account of adequately in the framework of an overall analysis.

Types of development programme: A development programme is an analysis which provides a basis for designing and carrying out development policy. There is, however, no sharp distinction between programming and policy making, since each influences the other. The main function of a programme is to make different policies consistent with each other. Ideally, it should go further and help to select the best policies and the best means of carrying them out. The decision to make a development programme does not constitute an endorsement of increased government intervention, or of any other particular set of policy instruments, therefore.

The nature of the analysis contained in a development programme is determined in part by the information available and in part by the instruments which are being considered. For simplicity, three general types can be distinguished, which I will call aggregate programmes, sector programmes, and overall programmes.

Aggregate programmes consist mainly of national accounts analyses and projections of other magnitudes such as industrial production, labour force, average productivity, etc. These projections are often combined with a more detailed analysis of certain aspects of the economy, such as the balance of payments, the sources of government revenue, etc.

Aggregate programmes provide a fairly adequate basis for the use of general policy instruments, but they do not furnish a check on the consistency of the results in specific sectors nor on the balance of payments. They are more likely to be adequate when the composition of production and consumption does not change too much as income increases, and when the market mechanism works well in directing investment and production decisions.

Sector programmes are analyses of the demands and investment prospects in individual branches of production. Their main function is to determine the relative priority of investments within the sector. Investment programmes for the whole economy (or for all resources controlled by the government) are sometimes constructed by merely adding up the high priority projects in each sector.

The sector approach is generally recognized to be inadequate as a basis for development policy because it does not provide a test of the consistency of the decisions made in each sector, nor a way of comparing high priority projects in one sector with those in another. It has nevertheless been the principal basis for development policy in Latin America and in most under-developed countries until quite recently. The defects in the sector approach are less serious in primary-producing economies than in those which have reached a high degree of industrialization and hence have a greater amount of inter-dependence among the various sectors.

Overall programmes combine the elements of aggregate programmes and sector programmes in varying degrees. The analysis may start from overall projections or from sector analyses, but in the final result they must be reconciled. It is only by some check of this kind that the consistency of the simpler models used in the two partial approaches can be tested.

The need for overall development programmes is most acute when large structural changes are required to establish or restore a process of balanced growth. Large balance of payments deficits, unemployment, bottlenecks in overall facilities, and lack of growth may each be evidence of such conditions. These conditions may of course be merely symptoms of an excess or deficiency of total demand, and the diagnosis of structural disequilibrium must try to identify the problems which would exist if inflationary (or, less often, deflationary) forces were offset. The design of policy in such circumstances is likely to call for an overall analysis, however, whether the policy measures selected are general or specific in nature.
We have repeatedly seen that in reality markets deviate from the ideal conditions of perfect competition. Because of market failure, market prices are unlikely to represent appropriate substitution ratios and an adjustment may be necessary before price is a true social coefficient of choice. As discussed in chapter 3, market failure can arise from existence of monopoly elements, lack of information, and externalities.

In situations of market failure, remedial policies are then necessary to remove the market imperfections or to internalize the externalities. A major task of any government is to devise remedial policies that will correct private market signals when necessary and improve the allocation of resources. Stated in formal terms, the objective should be to allocate resources in such a way that resources are devoted to an activity up to the point that the expected marginal social benefit equals the marginal social benefit in an alternative use (the forgone marginal social benefit is the marginal social cost). But to translate this formal condition into practice, a government may have to resort to remedial policy measures such as taxation, subsidization, regulation, or possibly establishment of a public enterprise.

Because of their prominence in developing countries, we gave special attention to public enterprises. Chapter 4 emphasized that the determination of public prices for public products should proceed beyond merely financial consideration to the efficiency and social dimensions of pricing.

Our earlier discussions of the elements of the price system, functions of prices, properties of efficiency, and market failure should help in the understanding of the efficiency and social dimensions of public pricing.

Our examination of public pricing ended with the recognition that pricing policy is closely related to investment policy. And this brings us to project appraisal—a central task for the government of every developing country.

**Toward Project Appraisal**

Some rule is necessary in deciding whether to accept or reject a project. In the absence of competitive markets in equilibrium, or when externalities are present, any evaluation of a proposed project must proceed to an adjustment of market prices before it can be determined whether the project should be accepted or not. The method of project appraisal must adjust the financial accountancy or "book" costs in order to reflect the "true" value of foreign exchange, labor, cost of capital, and so on. As a correction to market prices, accounting prices or shadow prices are needed to determine the least-cost solution for public enterprise investment programs.

The method of project analysis constitutes an entire subject in its own right. Selections 5.3, 5.4, and 5.5 provide only a preliminary introduction to the rationale and technique of project analysis. Although this book does not proceed further into the details of project analysis, it aims to provide a background that will make project analysis more meaningful. For project analysis is not simply a technique to be applied mechanically, but an approach to be interpreted. To aid in that interpretation, this book has established some principles of pricing that constitute the wider context in which project analysis must be practiced.
5.3 Cost-Benefit Analysis in Developing Countries

Projects are the building blocks of plans. In fact, we may define a project as the smallest unit of activity—from a technical or economic point of view—that can be undertaken. Such a unit of activity may be further defined as the coordinated use of scarce resources for the production of goods or services that increase national welfare relative to the situation without the project. There are several aspects to this definition that need elaboration. First, the word coordinated is used to indicate that some kind of organization—a private enterprise, a government department, a state corporation, a joint venture, or some other body—is necessary to operate the project. Second, we speak of a project only if its objective is to produce goods or services that increase national welfare. For instance, building a factory does not in itself constitute a project; since the factory should eventually produce goods or services, the project, in this case, consists not only of the investment in the factory but also of the factory’s operations, in the most general sense of the word, during its economic life. Third, the word investment is intentionally not used in the definition so that also such activities as research and extension can be considered projects. Finally, the definition implies that, somehow, national welfare should be maximized.

The theory of cost-benefit analysis provides criteria for the design and implementation of individual projects and for the selection of a series of projects to arrive at a country’s development program. The theory borrows substantially from both micro- and macroeconomic analysis in that the project must be considered both as an individual entity and also within the broader perspective of the national economy. An essential element of the theory is that it focuses explicitly upon the objective of increasing national welfare. What then is national welfare?

Often it is believed that national welfare can be maximized by maximizing profits. But if economic costs and benefits differ from financial costs and benefits, then profits cannot be taken to be the right measure for investment decisionmaking from a national point of view. In many cases there may be fundamental differences between a profitability analysis that uses market price data and a cost-benefit analysis that uses imputed values—the real economic values of costs and benefits. We may illustrate the issue with a number of examples.

First, suppose that a crop failure has increased the price of rice. With an inelastic demand curve, revenues and private profits will then be larger than they were before the crop failure. The use of the profit criterion leads thus to the ridiculous conclusion that a misharvest is better than an abundant crop. In reality, of course, the crop failure has reduced national welfare. In cost-benefit analysis this loss is measured by evaluating the loss in welfare of the consumers.

A second deficiency in the use of market prices arises when the factor of production in question is not fully employed. Assume, for instance, that there is unemployment. Then, the cost of hiring an unemployed laborer is obviously not measured by the wage cost, since the cost of employment is much greater than any marginal contribution to the production the laborer may have made in his unemployed situation.

Third, many countries employ systems of exchange control and licensing. In such cases the price of an imported good valued at the official exchange rate may well be a serious underestimate of the scarcity value of the import. The same may also apply in such cases to exports.

Fourth, inflation—a phenomenon almost pervasive in the present-day circumstances—can create serious distortions. Many costs and benefits may have no relation whatsoever to the real scarcity values of costs and benefits.

Fifth, a private profitability analysis pays often no attention to technological externalities, that is, to the effects of the contemplated investment on such things as water or air pollution or the environment in general. In a cost-benefit analysis such harmful effects must be taken into account.

Finally, as a last point it may be mentioned that a profitability analysis from a private point of view cannot capture considerations dealing with the distribution of income between persons and over time. A dollar accruing to a rich man is generally believed to be less valuable socially than a dollar accruing to a poor man. Hence, an increase in national income going to poor income groups would increase welfare more than if the income were to accumulate in the hands of the rich. Similar considerations may apply as regards the distribution of income between generations.

While a profitability analysis, as for instance undertaken by a firm, analyzes costs and benefits from a private point of view on the basis of market prices, the essence of cost-benefit analysis is that the costs and benefits of a program or project are measured on the basis of accounting prices or as they are sometimes called, shadow prices. These prices reflect the real scarcity values of the factor of production and of the

goods produced and also the indirect effects the project may have on national welfare.

Cost-benefit analysis should in principle be applicable to all kinds of projects. There are, however, projects whose benefits are difficult to quantify, such as those dealing with population control, health, education, and defense. Cost-benefit analysis has therefore developed mainly in the direction of projects with tangible benefits, such as those concerning agriculture, industry, transport, and other infrastructure projects.

National Planning and Cost-Benefit Analysis

In principle, the theory of cost-benefit analysis considers only those variables that are relevant to the project under consideration. It is possible, therefore, that one could imagine that all these variables should be determined by the project preparation agency. This would be a mistake because several project variables can be determined only at the macro level. For instance, the demand for a product depends inter alia on population and income growth; the supply situation or a project's input may depend on the growth path of the economy; the shadow price of foreign exchange and the discount rate to be used for project planning are clearly national parameters; and so on. If every project evaluator were to determine individually the national parameters, there would be quite a duplication of effort. Moreover, in all probability, one would end up with divergent estimates and misinvestments because the government departments and public corporations that prepare the various investment projects would not have sufficient information to estimate the national parameters meaningfully. It is the task of the central authority—the Central Planning Unit or the Treasury—to review the projects prepared and proposed by the different departments and corporations and to determine after consultation with them what the final set of projects will be. But it is this final set that determines how the economy will develop, so that only the Center can really determine what the values of the national parameters are. Furthermore, it may so happen that the available investment funds will not be sufficient to undertake all the proposed projects. In such a case, it is the Center's task to set the cut-off rate of return for the individual projects at such a level that the number of acceptable projects will just exhaust the available budget. This rate also can be determined only at the Center.

A national plan consists normally of sectoral plans so that the possibilities within a sector can be reviewed and misinvestments prevented. Suppose, for instance, that every region in a country plans to undertake a sugar project. Some sectoral review is then required to prevent the likely overinvestment. Also, sectoral plans will be necessary when different types of projects within a sector are interconnected. For instance, in the transport sector, a port project may necessitate improvement of road and rail connections with the hinterland, or the improvement of a certain road may necessitate a feeder road improvement program, and so on. Furthermore, it is possible that a shortfall of investments in the transportation sector may hold back development elsewhere in the economy. Clearly, if the sector is so important, some long-term planning is necessary. Another example would be the following. Assume that a large power resource—hydroelectric, gas or oil—can be developed at low cost. Then it may be in the country's interest to set up a long-term plan for the promotion of energy-intensive industries. In general then, sectoral planning will be necessary where overinvestments may occur, where important interrelations within the sector exist, or where the sector will have an important impact on the rest of the economy.

In addition to sectoral planning, plans may be required for regions or income groups. As regards regions, important interconnections may exist between agriculture, industry, transport, labor, etc., which make such plans mandatory. As regards income groups, it is now more and more accepted that for certain groups some form of planning is required. Assume, for instance, that the income level of a certain income group, say, smallholder farmers, is expected to remain depressed. Assume further that the government has decided to remedy the situation by means of an investment program geared toward raising the income level of the group; then also a series of projects spread out over time and interconnected, i.e., a plan, will be necessary.

Since most sectoral plans will have a substantial impact on other sectors and plans for regions or income groups will cover several sectors, the drawing up of the plans should in most cases be done jointly by the Center and the concerned government departments. The Center thus plays a crucial role. It is responsible for the macro-economic plan, it has to assure itself that the intermediate level plans relate properly to the overall plan, and it has to decide which set of individual projects will be implemented. In all of this it needs project data for the calculation of the national parameters but, at the same time, the lower levels can only provide the project data after they have received the national parameters.

Although such interactions might appear to involve an insolvable circularity, the process, if it is appropriately handled, would sharpen the understanding of the central authority as regards the desired development path of the economy. In a well-run administration the interaction might take place as follows. The central authority would make, with the help of programming models, tentative projections of development objectives and national param-
eters, which it would submit to the various government departments. These would use the data in the preparation of projects within their jurisdiction and submit to the Center a list of projects together with feasibility studies. The Center would review how the proposals would fit in with the initial development strategy, calculate new national parameters, including tentative cut-off rates of return in case of a shortage of available funds, and submit them to the departments, which would in turn revise their plans and submit them to the Center, and so on. As a result of these interactions, it can be expected that a well-formulated national plan, including intermediate level plans, and a set of well-prepared projects would emerge.

It must be remarked that the above sketch presents the target toward which the planning process should strive. In many countries we still find that the concerned government departments have no planning unit or that the interaction between the Center and the Department is still so rudimentary that the latter has little notion of the values of the national parameters. It is encouraging to see, however, that all this is gradually changing for the better.

In the past, many governments relied mainly on macro-economic plans for their investment strategies. Often these plans were based on estimates of simple sectoral capital-output ratios and, as a result, serious misinvestments occurred. Take, for instance, the commonly used incremental output to capital ratio $\Delta Y/\Delta K$. The theory underlying the use of this indicator is that investments should be stepped up in those sectors where the ratio has a high value. In other words, investments in a sector where $\Delta Y/\Delta K$ has a value of, say, $\frac{1}{3}$, which is taken to mean that an additional dollar invested has a return of $33\%$, is assumed to provide more benefits than investment in a sector where $\Delta Y/\Delta K$ has a value, say, of $\frac{1}{5}$, i.e., a $20\%$ rate of return. The deficiencies of this approach are clear. First, the benefits will not be produced immediately so that the actual rate of return will be much lower. For instance, if $\Delta Y$ were to materialize after four years and then to be produced in perpetuity, then the rate of return $\gamma$ should be calculated from $\Delta Y (\gamma_1 + \gamma)^4 = \Delta K$. Thus if $\Delta Y/\Delta K$ is $\frac{1}{3}$, then the rate of return $\gamma$ would be $16\%$ instead of $33\%$. But also this would be an overestimate since projects do not have infinite lives. If one wants to work with a rate of return, then one should draw up the stream of costs and benefits of the investment and calculate the internal rate of return of these streams. Second, the ratio does not reflect accurately the costs that must be incurred to produce the additional benefits. The ratio states that incremental output will be produced by additional investment. Obviously, however, this cannot be correct because output is not only a function of capital, but also of labor. The formula assumes that labor has a zero cost and this is, of course, a very unrealistic assumption.

Our criticism of the output to capital ratios should not be taken to mean that the ratios have no role to play in the drawing up of investment strategies. They do, but the resource allocations emerging from the models, even if more refined ratios are used, are still so crude that the possibility of serious mistakes cannot be excluded. Cost-benefit analysis of individual projects has a crucial role to fulfill in that it can reduce these errors. As it is based on detailed investigations of the economic viabilities of individual projects, it can ensure that only projects with high priority will be selected. Similarly, as regards intermediate level plans, cost-benefit analysis can be very useful. If a certain sector's initial set of projects includes some with low rates of return then it would be in the national interest to transfer part of the initial budget allocation for that sector to a sector where projects have higher rates of return. Cost-benefit analysis thus not only increases the likelihood that the investment program will be successful, it leads also to an optimal program. Fortunately, it is becoming increasingly accepted that the basis on which macro plans as well as intermediate level plans rest are well thought out individual project proposals. Cost-benefit analysis is the tool that provides this basis.

The Derivation of the Shadow Prices

In many economies market prices do not reflect the real scarcity values of the goods and services that are being produced. Monopolies, decreasing cost industries, taxes, externalities, foreign exchange and capital scarcities, unemployment and underemployment, government fixation of prices and wages, inflation, and so on are common in most of the developed as well as the developing countries.

Is it at all possible, in view of the many difficulties, to determine all the relevant shadow prices? It has been suggested that they should be derived from general programming models. Such models can generate Lagrangean multipliers that represent in economic terms the shadow prices, which, given the constraints incorporated in the model, will result in an allocation of resources that will satisfy the postulated objective function. Such models should, in principle, be able to produce the real scarcity prices of the goods and services. However, although the models can provide valuable insights concerning the structural relations that exist in an economy, we seriously question on practical grounds whether all the shadow prices obtained from the models can be used for operational work. First, the models are still highly aggregated so that most of the duals they generate—the shadow prices—are extremely crude. Second, it must seriously be doubted whether the models can
really depict the real world situation. It is not only that there are many distortions but also, to formulate the models, all cost and demand functions should be known. Obviously it is impossible to collect all these data. We feel, therefore, that in addition to the models approach a more practical approach must be followed.

In fact, this approach—the opportunity cost doctrine—has existed for a long time. Consider an economy where only two goods, X and Y, are produced. Then the calculation of the costs of an output expansion would not pose a difficult problem. For instance, if the production of X is to be increased, the cost of producing the additional quantity of X is to be found by measuring the value of the Y goods that the community will have to give up in order to increase the production of X. Analysis of the production and demand functions for goods X and Y should readily provide the required data. The situation is more complicated when a multiplicity of goods is being produced, since it will obviously be impossible to analyze the production functions and demand curves of all the different goods. The opportunity cost doctrine therefore takes as a starting point the inputs which X uses rather than the displaced Y goods and defines the costs of these inputs as the returns that they would earn in the next best alternative elsewhere.

The opportunity cost approach is necessarily a detailed approach. The project analyst must investigate from where the resources for a project will be withdrawn and what their values are in those uses. As long as resources with low valued uses can be transferred to higher valued uses, the change is beneficial. This, in a nutshell, is what the theory of project planning is all about.

Without any claim for comprehensiveness, we may now illustrate how the shadow prices can be found in practice. When a project needs a certain input, then it is likely that to some extent demand elsewhere will be curtailed as well as that some additional quantities will be produced to satisfy the additional demand generated by the project. Thus the opportunity cost of the input consists of the weighted average of the value forgone in the alternative use and the resource cost of the additional production, the weights being the fractions of demand displaced and supply induced to additional demand. What will happen in practice depends on the shape of the demand and cost functions and theoretically the analyst needs therefore to investigate these functions in detail. In real life, however, it often suffices to make some rough estimates as the project may not be sensitive to the value of the input in question.

Foreign exchange is for many countries such a scarce resource that it can be treated as a separate production factor. To determine the shadow price of foreign exchange, we apply the same analysis as for an input. We should thus consider how an additional dollar of foreign exchange can be obtained and then determine the real resource values in the forgone uses. In principle, there are two ways to obtain foreign exchange: curtail imports or increase exports.

In case of a curtailment of imports, the shadow price of foreign exchange can be found by comparing the domestic value of the import with its c.i.f. value converted at the official exchange rate. Hence, if there are no quantitative restrictions, this value will be higher than the official exchange value by the amount of the import duties. Thus if the official exchange rate is US$1.00 = Rs2.00 and the import duties are 30 percent, the domestic value of US$1.00 of foreign exchange is Rs2.60.

If exports are to be increased, then the resources otherwise used to produce home goods will be used to produce export goods. Hence, the shadow price of foreign exchange consists then of the resource value of the forgone home goods. If there are no restrictions on exports, the level of the export duties or subsidies will provide us with a good indicator of the domestic value. For instance, if export duties average 10 percent of f.o.b. value of Rs0.20 per US$1 exported, then this means that exports are Rs0.20 more expensive than the home goods that could be produced with the same resources. Hence, the domestic value of the exports and the shadow price of foreign exchange would then be Rs1.80 = US$1.00.

As in all cases where a resource is used, there are thus two shadow prices corresponding to whether the resource comes from a displaced use or is additionally produced. As in reality there will be some combination of the two possibilities, the correct shadow price will be a weighted average of the two shadow prices. Now, with respect to foreign exchange it is, of course, extremely difficult to determine by how much exports will be increased and imports reduced if additional foreign exchange is needed. A shortcut method is to work with a normal average of the two shadow prices and, in our example, this would be the average of 2.60 and 1.80, so that the base estimate can be set at Rs2.20. Further refinements can be made by considering whether one or the other possibility of earning foreign exchange is the more likely. For instance, if in our country import curtailment is easier than increasing exports, then the foreign exchange shadow rate can be estimated at somewhere between Rs2.20 and Rs2.60. On the other hand, if it is easier to increase exports, then the shadow rate could be set at somewhere between Rs1.80 and Rs2.20. Whatever value one chooses, in all cases sensitivity tests should be applied.

As regards the shadow price of labor, many developing countries have still a shortage of skilled labor so that market wages can often be taken to represent the real scarcity values of this production factor. In several developing countries, however, un-
skilled labor is available in abundant supply. The determination of its shadow price is distinct from all other production factors because of the fact that the laborer's services are tied to the laborer. This means that, if a laborer is withdrawn from existing employment to work in a new job, then not only the forgone product of the laborer but also the disutility of his extra effort should be taken into account to determine the shadow price. Thus, even if the forgone product of a hired unskilled worker is negligible, the shadow price must be set at the price which will induce him to work in the new job.

In many cities in the developing countries, we see that the influx of rural workers is greater than the job openings in the formal well-paid sector. In such cases, more than one person migrates to the cities when one person is hired. What this means is that in order to find the shadow price of labor, we must add to the opportunity cost of the hired laborer the opportunity cost of the workers who have migrated, but who cannot find a job in the formal sector. The shadow price of labor in the cities may thus be a multiple of the rural shadow price of labor. It is, of course, very difficult to determine precisely what the exact migration function is, as often no reliable data will be available. The best one can do in such cases is to assume that the shadow price of urban labor lies somewhere between the institutional and the rural wage rate adjusted for cost-of-living differences. The use of both values will show whether the project is worth undertaking. Two criteria have become common: the net present value and the internal rate of return criterion.

The Calculation Techniques

After having imputed the correct scarcity values to all the inputs and outputs of a project, including negative values for harmful effects on the environment, the task at hand is to determine whether the project is worth undertaking. Two criteria have become common: the net present value and the internal rate of return criterion.

Under the net present value criterion all the costs and benefits of the project are discounted at the opportunity cost of capital to present values, and the project is considered worthwhile if it has a positive net present value. The criterion amounts thus to the calculation of the present value of the surplus the project generates over and above the opportunity cost of capital.

The internal rate of return criterion consists of calculating by trial and error the discount rate at which a project has a net present value of zero—the project's internal rate of return—and accepting the project only if its internal rate of return is larger than the opportunity cost of capital.

Is the internal rate of return criterion a theoretically correct criterion? The answer is negative and the criterion may lead to wrong results if projects are to be ranked by priority. The reason for this is that the internal rate of return criterion assumes incorrectly that the benefits of the project under scrutiny will be reinvested at the internal rate of return instead of at the opportunity cost of capital as assumed under the net present value criterion. It is intuitively clear that the two criteria will therefore result in a different ranking of projects. If budgetary constraints play a role, so that a choice must be made out of a series of projects, then the only correct criterion is the net present value criterion. The way to proceed then is to calculate for each project the present value of the surplus it generates per dollar of current investments,
and to choose those projects that have the highest surplus ratios.

The problem with the net present value criterion is, however, that the opportunity cost of capital, which serves as the discount rate, is difficult to estimate. One may attempt to calculate net present values by using a minimum and a maximum value for the opportunity cost of capital, but where one can only estimate that the opportunity cost of capital lies between some very extreme values, this procedure is not very meaningful. In such a case, the internal rate of return method may be used to establish a tentative ranking order of projects. As projects with high internal rates of return will be accepted in any case, one can then spend extra time on projects with low rates of return—those at the margin of acceptability—in order to sharpen the accept-reject decision. This may be done for instance by calculating probability distributions of the costs and benefits. We feel, therefore, that the internal rate of return is an important tool in practical work despite its theoretical limitations.

5.4 Some Preliminary Theory

The purpose of this paper is to compare and critically evaluate various alternative project selection procedures which have been put forward, particularly for application in less developed countries. It is meant primarily for operationally occupied economists who may be confused by the various brand names, as well as the esoteric and highly charged claims and counter-claims made by the proponents of the different methods, and who, moreover, may want to learn how the different methods fit in with economic theory as well as their own immediate practical preoccupations.

Some Preliminary Theory

It is appropriate to begin by stating the obvious: cost-benefit analysis is undoubtedly the most used, and arguably the most useful, form of applied welfare economics. Its theoretical basis as well as its limitations are therefore necessarily those of its parent, theoretical welfare economics. This paper is not for those who deny any practical use for theory, but for those who whilst recognizing the limitations of theoretical welfare economics nevertheless feel that in our present state of knowledge it provides the only basis for making an economic assessment of investment plans and proposals.

The purpose of any project selection procedure must be to provide a decision rule for accepting or rejecting a project. The net present value (NPV) or the internal rate of return (IRR) of the project are the indices usually used. Our chief concern in this paper will be with first, what should be included in the time stream of benefits and costs; secondly, what are the relevant values of the various cost-benefit components; and thirdly, how the discount rate (or rates) needed for determining the NPV, or the cut-off IRR at which projects are accepted, should be chosen. Most of the differences in the alternative procedures relate to apparently differing prescriptions in these three respects.

It will be repeatedly emphasized in this paper that any substantive differences among the alternative procedures are in large part dependent upon differing assumptions about the relevant aspects of the economic environment in which the investment decisions are being made. One of the basic purposes of this paper will be to demonstrate that, in principle, most of the suggested procedures are equivalent, if the same assumptions are made about the economic environment, though naturally there are differences in emphasis as to which set of assumptions is more relevant for LDC's in general, and more importantly in the practical problems of estimating the relevant values to be included in the NPV/IRR index, with accuracy and ease.

The reason why in principle most of the methods are equivalent, given the same basic assumptions about the economic environment, is their common lineage—theoretical welfare economics. One of its basic results is that in a perfectly competitive economy (with no uncertainty about future tastes and technology), allocation of resources on the basis of market prices of goods and factors (for which markets exist) would result in Pareto optimality for a given income distribution. Market prices of goods and factors would equate and equal the marginal social cost (MSC) of producing and the marginal social value (MSV) of

3. Pareto optimality necessitates that for a given distribution of income:

(i) the marginal rates of transformation in production of different commodities are equal to their marginal rates of substitution in consumption,

(ii) the marginal rates of substitution between any pair of factors are the same in all the industries in which they are used,

(iii) the marginal rates of substitution of any pair of commodities is the same for all individuals consuming both goods.

Given that the above conditions hold, a Pareto optimum will exist, such that for the given income distribution it will not be possible to make one person better off without making someone else worse off. Treating the same physical commodity at different dates as many different commodities, equivalent intertemporal marginal equivalences for an efficient intertemporal program can be derived.

using the relevant goods/factors. For a truly marginal investment project (in the sense that it does not alter the MSV and MSC's of its output and inputs it uses as a result of its operation), the values of the output and inputs at market prices would provide the correct values to be used in determining the net present value of the project. Market prices would be the "shadow" prices to be used in project selection.

If the investment project being considered is not marginal (or if there are externalities), and does affect the MSV and MSC's of its output and inputs, then the relevant measures of the social benefits and costs of the project will be the change in the consumers' and producers' surpluses caused by the project. This, in principle, will be the procedure recommended by all the project selection procedures we shall consider. In the case of the perfectly competitive model, valuation of the changes in producers' and consumers' surpluses, at market prices, will provide the correct indication of the net social benefits of the project.

To the extent, however, that the perfectly competitive paradigm does not hold—for example due to the existence of monopolies, taxes and subsidies, externalities, and/or increasing returns—market prices will no longer indicate the social costs and benefits of using and producing different commodities. The social cost to be included in the NPV/IRR index of social profitability, properly defined, will still be the marginal social cost of the various inputs used, and the social benefit will be the marginal social value of the output produced. However, the breakdown of the perfectly competitive assumptions results in market prices no longer equating and equaling the MSC and MSV of the relevant commodities. The market price will now equal either the MSV or MSC—and in some cases of rationing may not equal either. The problem then is to adjust the market price to obtain the relevant "shadow" prices, which are therefore generally needed in investment appraisal because of the divergence between the MSC and MSV of the relevant commodities.

If neutral fiscal devices (lump-sum taxes and subsidies) are feasible, then a full Pareto optimum could still be achieved if the government eliminates the divergence between MSC and MSV by suitably correcting tax-subsidy measures, thereby restoring the equivalence of MSC and MSV with the market price of the commodity. However, for obvious reasons it will not be possible, in most cases, to cure the divergence in this manner. In that case, the divergence between the MSC and MSV of the commodity may have to be taken as a datum (or a constraint) and the "shadow" prices corresponding to this constrained (or "second-best") welfare optimum will need to be computed. A large number, if not most, of the shadow prices which we shall consider are of this "second-best" kind.

Second, even if the government can eliminate the divergence between MSC's and MSV's by suitable tax-subsidy policy, it may take time for the divergence to disappear. Then current market prices will not equate the MSC and MSV of the relevant commodities, but it is expected that future market prices will. As investment takes time and its effects are extended into the future, it is clearly the MSC's and MSV's of the relevant inputs/outputs appropriately dated which will be relevant in working out the project's social profitability. If it appears likely that in the future an existing divergence between MSC and MSV will be corrected, the appropriately dated price which reflects the social cost/benefit of the project will not be the current market price, nor the current MSC and MSV of the commodity, but rather the "equilibrium" price which is expected to prevail in the future. In this sense, even when an economy is moving towards an optimal set of market prices, from a distorted current set, it may be necessary to use "shadow" prices corresponding to the future optimal market prices, rather than the current market or shadow prices for pricing inputs and outputs which form the time stream of benefits and costs of the investment project.

Third, even for a perfectly competitive economy, there will be different Pareto optima associated with different income distributions. Judging between these different Pareto optima will necessarily involve normative judgments about the desirability of particular income distributions. Even if agreement can be reached on the desired income distribution, there will still be the problem of legislating this "optimal" distribution. Again if neutral fiscal devices in the form of lump-sum taxes and subsidies are feasible, the government would be able to achieve a Pareto optimum with the optimal distribution of income. If however, as is more likely, neutral fiscal instruments are not available, then the distributional effects of investment projects will also have to be computed, and judged against and along with their purely "pro-

4. The net benefits being discounted at the optimal discount rate which equates the marginal rate of transformation (mrt) in production of present into future consumption, to its marginal rate of indifferent substitution (mrs) in consumption, determined in a perfect market for intertemporal consumption.

5. It being noted that investment projects affect both the intratemporal as well as the intertemporal distribution of income; the former by the distribution of their net benefits amongst contemporaries at a point in time, and the latter by the distribution of net benefits as between generations, over a period of time.
duction" or "efficiency" effects. These problems open up other areas where there may possibly be conflicting judgments, and hence prescriptions for project selection procedures.

Practical Problems

These theoretical problems are compounded by practical ones. First, even though there may be agreement about the nature of the correct prices to be used in project selection, there may, nevertheless, be disagreement as to whether or not the existing divergences between MSV's and MSC's which effect these prices will continue into the future or whether they will change. Depending on what assumption is made about the future course of the economy, the "second-best" or "first-best" shadow price will be the relevant one to choose. In a sense, this is an empirical question; but to the extent that future government policies are normally unknown, the element of judgment involved in deciding which of these alternative assumptions is relevant, when considering existing distortions in commodity and factor markets, will be of paramount importance in deciding which is the correct "shadow" price to use. Hence it is important to remember that differing prescriptions on alternative evaluation procedures will most often be due to differing implicit assumptions about the current and, more importantly, the future economic environment.

Second, though we have been discussing the evaluation of a particular project and the social valuation of its inputs and outputs in what may appear to be a partial equilibrium framework, in principle, any proper investment criteria must take account of the total (direct and indirect) or what are termed the general equilibrium effects of the investment project. Thus for instance if an industrial project employs some seemingly underemployed labor in the urban sector, the ultimate effects via the impact on rural-some seemingly underemployed labor in the urban sector, the ultimate effects via the impact on rural-urban migration could be a significant change in total output of the economy. The shadow wage rate will then in this case have to incorporate both the direct and indirect (via migration) effects of increasing industrial employment. The MSC's and MSV's which are taken as the "shadow" prices in determining the social profitability of the investment project, must therefore be the general equilibrium "shadow prices." This might appear to be an impossible task, but the relative merits of alternative investment appraisal procedures will depend upon their success in taking account of the general equilibrium effects of projects, which will in turn, if the procedures are to be practical, necessitate making certain simplifying assumptions about the economic environment. Once again, these assumptions, though empirical in nature, require judgment, and hence there can be disputes as to whether or not the simplifying assumptions are "realistic" or relevant or both.

For all the above reasons, even though all the procedures we will consider start from the same theoretical foundations, and hence are identical if equivalent assumptions are made, they may nevertheless differ to the extent that, in practice, they emphasize one set of assumptions about the economic environment rather than another. Hence, the continuing charges and countercharges that a particular procedure has ignored or assumed away an important aspect of reality, and is hence invalid; as well as the impression conveyed to neutral observers of shadow boxing on the part of different protagonists, and bafflement at the conflicting claims and counterclaims that are made for different procedures. This, however, does not imply that in practice certain procedures are not more general and easier to apply than others. However, it may be more important to begin by realizing that the similarities amongst the procedures are far greater than the differences.

5.5 Nature of Project Analysis

The basic economic problem facing all countries is that of allocating inherently limited resources (such as labor, capital, land, and other natural resources, as well as foreign exchange) to a variety of different uses (such as current production of consumer goods and public services as against investment in infrastructure, industry, agriculture, or other sectors of the economy) in such a way that the net benefit to society is as large as possible. Given the limitation of resources, choices must be made among the competing uses, and project analysis is one method of evaluating alternatives in a convenient and comprehensible fashion. In essence, project analysis assesses the benefits and costs of a project and reduces them to a common yardstick. If benefits exceed costs, with both measured by the common yardstick, the project is acceptable; if not, the project should be rejected.

In assessing the merits of different projects, the objectives of any particular society clearly must be taken into account. That is, project costs and benefits must be measured against the extent to which

6. The second-best shadow price is that associated with continuing divergences, the first-best, that with no divergence, between MSV and MSC.
they detract from, or contribute to, achievement of that society's objectives. At a general level, nearly all countries may be assumed to have two primary and simultaneous—if not always equally valued—objectives: to increase total national income, the growth objective, and to improve the distribution of national income, the equity objective. In general, therefore, projects should be assessed in relation to their net contribution to both of these objectives, but this has not always been the practice of the World Bank or of other lending institutions.

**Traditional Practice**

Until recently, traditional methods of appraisal have emphasized the growth objective, often to the detriment, if not the virtual exclusion, of the equity objective. This has been justified on the grounds that governments have available to them a diversity of fiscal devices that can be used to redistribute project-generated income in any desired direction. It was argued that project analysis need consider only the growth objective, since this would ensure that the available resources yielded the maximum increment in total national income; the equity objective could then be served by a program of taxes or subsidies that would bring about the desired redistribution of that maximum increment in national income.

At the practical level, the policy concern with growth was interpreted to mean that projects should be selected in the light of their contribution to the maximization of total undifferentiated national income. This appears at first blush to be a faithful interpretation of the policymaker's concern with growth, but it is strictly correct only if it can be assumed that at the margin all units of project-generated income—that is, a unit of either investment or consumption—make the same contribution to growth. This assumption was generally accepted, either explicitly or implicitly, in traditional practice, as well as in most of the earlier theoretical literature on project analysis. As a result, when theorists attempted to derive, and practitioners to estimate, shadow prices that would reflect the true value of inputs and outputs to society better than market prices, they assumed that at the margin all units of income were equally valuable from the growth point of view and ignored the equity objective. Such shadow prices are here referred to as "efficiency prices."

**Recent Innovation**

More recently, however, it has been argued that the operational assumption that all units of income make the same contribution to growth may be untenable. For example, in an economy in which the level of national investment is below what is required to secure the desired level of growth, investment may be considered more valuable than consumption. If this argument is accepted, the successful pursuit of the growth objective requires that the distributional effect of the project on consumption and investment be included in the overall assessment of the project's worth, and that project-generated income that leads to investment be assigned a higher value than that which leads to consumption. In this fashion, use of investment resources will be biased in favor of projects that generate more investment, thereby raising national investment toward the desired level.

The validity of this new argument turns on the extent to which the government is free to determine the desired level of investment by means of the traditional instruments of fiscal and monetary policy. Clearly, if the government controls the level of investment in such a manner that at the margin society is indifferent between a unit of investment and a unit of consumption—that is, a unit of either would make the same contribution to welfare—there is no need to differentiate between project-generated income that accrues in the form of investment and in the form of consumption. It can be argued, however, that there is a diversity of social, administrative, and political constraints, especially in developing countries, that may inherently limit the government's ability to increase savings by means of monetary and fiscal policy. And if these tools of general economic policy cannot successfully break the diverse constraints, other policy instruments, including the selection of projects, can and perhaps should be used to achieve the desired goal.

Theorists and practitioners therefore turned their attention to the derivation and estimation of shadow prices that recognized a suboptimal rate of investment. But once there was a recognition of the constraints on the government's ability to secure the desired distribution of income between investment and consumption, it was but a short step to the realization that the separation of the growth and equity objectives may not be justified: that is, that the government's ability to redistribute income in general may be limited. On this basis it was concluded that project analysts should investigate the impact of projects not only on the distribution of income between consumption and investment but also on the distribution of income between the rich and the poor. Shadow prices that include both these distributional aspects are here described as "social prices."

This study may be viewed in part as an attempt to provide an up-to-date statement of these ideas and to summarize the main developments in the recent literature as they relate to the traditional practice of the World Bank and other lending institutions. In particular, it is shown that the traditional method of project analysis represents a special case.
of the more general method recommended here. The text attempts to clarify the conditions under which the traditional method is appropriate and to define more clearly the nature of some efficiency prices (for example, the shadow exchange rate) and how they can be estimated. This part of the discussion, however, does not basically alter the standard practice followed thus far with respect to calculations of shadow prices and rates of return; it seeks simply to systematicize existing practice into a more consistent methodology.

In addition, and perhaps more important, the study seeks to provide more specific guidance than is available elsewhere on the mechanics of incorporating considerations of income distribution into the traditional calculation of rates of return. Basically, this involves attaching suitable weight, to be determined by the appropriate decisionmaker, to the benefits from the project that accrue in different forms (consumption or investment) and to different beneficiaries (rich or poor). Such weights might be assigned directly or, as described in the text, they may be derived from underlying notions of a welfare function. The most important issue, however, is not the technique for deriving the weights, which will undoubtedly be refined in due course; it is that, whatever weights are considered to reflect properly the relative value attached to benefits for various higher or lower income groups and to additional investment, these weights be consistently and systematically applied when evaluating the socioeconomic merits of a project. Only in this way will it be possible to allow in project decisions for the tradeoff between raising consumption levels of the poor and accelerating overall economic growth. Project evaluation procedures that neglect these aspects in their decisionmaking criterion are not responsive to recent concern with the distribution of the benefits of economic growth or to earlier concern with the achievement of growth. Evaluation procedures employed by the World Bank and other agencies, including national governments, should be consistent with general policy on these matters. This means, for the Bank and other external agencies, that the appropriate set of weights for project analysis must be worked out in cooperation with the client countries.

Possible Objections

Although the method outlined in this study has been widely accepted as, in principle, an important step forward in improving appraisal techniques and practices, doubts have been expressed with regard to (a) some aspects of the methodology, in particular, its appearance of spurious accuracy and dependence on dubious assumptions; (b) the possible lowering of rigorous appraisal standards; (c) its practical significance for project decisions; (d) its feasibility in regular operations of the World Bank and other agencies; and (e) its desirability, in view of the expected costs and benefits of the proposed approach. These objections and related issues are discussed below.

There is, of course, ample scope for further improving the methodology set forth in this book. The techniques for deriving and estimating shadow prices will undoubtedly evolve with further practical experience. Some elements of the analysis, particularly with respect to the distributional impact, are necessarily rather crude. In all cases, though, it is important, first, that all major considerations such as income distribution and fiscal constraints be included in the analysis (although in a crude fashion); and, second, that shortcuts such as those discussed in the text be adopted in full awareness of the approximation involved and the conditions in which they are justified.

The reader should not be misled by the apparent accuracy of the methodology. The equations describe concretely the specific relations underlying the analysis. They make clear what considerations were taken into account and what assumptions were made. This should not be taken to imply, however, that the postulated relations are accurate in the sense of providing a complete picture of reality. They are thought to give a good representation of the more significant features that should be taken into account in many cases, but, as the text tries to make clear, other factors and other relations may give a better approximation in some cases. Because both the nature of the relations among key factors and the magnitude of the parameters governing these relations are subject to substantial uncertainty, we do not claim great accuracy for the proposed method of analysis. At the same time, the cost estimates, demand projections, and price forecasts that are included in more conventional cost-benefit analysis are also beset by a large margin of error. The real point is not the degree of accuracy but the taking into account of all major factors bearing on a project decision as best one can rather than simply ignoring them.

Again, the simple welfare function that provides the central theme of the discussion of distributional aspects of project analysis can easily be criticized as oversimplistic. Others may prefer more complex functions that allow for a greater range and variety of judgments on equity and social justice. We believe, however, that many basic features of distributional problems can be satisfactorily clarified and reflected in project analysis by the simple welfare function used. Furthermore, more complex functions are likely to cause even greater problems of estimation.

The use of distribution weights to determine the social return on projects does not mean that appraisal
standards would be lowered. On the contrary, the methodology outlined in this book involves a more rigorous and systematic analysis than normally is carried out. The criterion for acceptability of a project is changed, but it is not more lenient. Some projects that otherwise would have been rejected will be acceptable because their distributional effects are given weight in the criterion; some otherwise acceptable projects will be rejected because of an adverse distributional impact.

Nor does the criterion neglect growth: the weight assigned to investment (and hence growth) is always positive and is determined jointly, on the basis of the same welfare function, with the weights attaching to benefits to richer and poorer people. Society's concern with growth can therefore be balanced against its concern with current consumption and equity in such a manner that neither concern is neglected.

The recommended approach should be more rigorous than current practice in another respect: it envisages a more systematic assessment and a more consistent application of the traditional efficiency prices, as well as of the recommended social prices incorporating distribution weights, than is customary in traditional analysis. With respect to both efficiency and social prices, it is important that public agencies, national or international, use the same estimates in the analysis of all projects in a particular country in order to provide greater assurance that shadow prices are estimated and applied in an objective and unbiased manner. This need for consistent use of shadow prices applies to all stages in the analysis of investment choices up to the final decision to proceed with the project.

In general, the significance of the proposed method of analysis is that it produces a systematic bias in project decisions favoring projects that benefit the poor rather than the rich and that result in higher savings and further growth rather than higher current consumption. Such an outcome represents not a distortion but is, since the extent of these biases is determined by the distributional weights, which are set in such a way that they reflect the fundamental socioeconomic objectives of the particular society. As a result, projects will tend to be selected or rejected with due regard to their impact on income distribution and growth. Although these factors may or may not be decisive for any particular project choice—depending on the distributional effects and the weights given to them—we expect this analysis to result in a pattern of investment that will differ significantly from the pattern that would emerge if distributional considerations were to continue to be ignored.

As in all project appraisal methods, the significance of this new methodology becomes greater the earlier it is applied in the project cycle. This underscores the importance of focusing economic analysis on the project at the time when its design is taking shape and when choices are still open, rather than when the project has become frozen and rejection may be difficult. But this is not an argument for or against any particular method of project analysis, including the one set forth here.

All countries, but particularly the developing countries, are faced with the basic economic problem of allocating limited resources such as labor at all levels of skill, management and administrative capacity, capital, land and other natural resources, and foreign exchange, to many different uses such as current production of consumer goods and public services or investment in infrastructure, industry, agriculture, education, and other sectors. These different uses of resources, however, are not the final aim of the allocative process; rather, they are the means by which an economy can marshal its resources in the pursuit of more fundamental objectives such as the removal of poverty, the promotion of growth, and the reduction of inequalities in income. Using limited resources in one direction (for example, investment in industry) reduces the resources available for use in another direction (investment in agriculture). Pursuit of one objective (better income distribution) may involve a sacrifice in other objectives (rapid growth).

Thus, there are clearly tradeoffs: a country can have more of some things and less of others, but not more of everything at once. A choice therefore has to be made among competing uses of resources based on the extent to which they help the country achieve its fundamental objectives. If a country consistently chooses allocations of resources that achieve most in terms of these objectives, it ensures that its limited resources are put to their best possible use.

Project analysis is a method of presenting this choice between competing uses of resources in a convenient and comprehensible fashion. In essence, project analysis assesses the benefits and costs of a project and reduces them to a common denominator. If benefits exceed costs—both expressed in terms of this common denominator—the project is acceptable: if not, the project should be rejected. As such, project analysis may appear divorced from both the fundamental objectives of the economy and the possible alternative uses of resources in other projects. The definition of benefits and costs, however, is such that these factors play an integral part in the decision to accept or reject. Benefits are defined relative to their opportunity cost, which is the benefit foregone by not using these resources in the best of the available alternative investments that
cannot be undertaken if the resources are used in the project. The forgone benefits are in turn defined relative to their effect on the fundamental objectives. By defining costs and benefits in this fashion we try to ensure that acceptance of a project implies that no alternative use of the resources consumed by this project would secure a better result from the perspective of the country's objectives.

Economic analysis of projects is similar in form to financial analysis in that both assess the profit of an investment. The concept of financial profit, however, is not the same as the social profit of economic analysis. The financial analysis of a project identifies the money profit accruing to the project-operating entity, whereas social profit measures the effect of the project on the fundamental objectives of the whole economy. These different concepts of profit are reflected in the different items considered to be costs and benefits and in their valuation. Thus, a money payment made by the project-operating entity for, say, wages is by definition a financial cost. But it will be an economic cost only to the extent that the use of labor in this project implies some sacrifice elsewhere in the economy with respect to output and other objectives of the country. Conversely, if the project has an economic cost in this sense that does not involve a corresponding money outflow from the project entity—for example, because of environmental effects or subsidies—this cost is not a financial cost. The two types of cost need not coincide. Economic costs may be larger or smaller than financial costs. Similar comments apply to economic and financial benefits. Economic costs and benefits are measured by "shadow prices," which may well differ from the market prices appropriate for financial costs and benefits.

Shadow prices are determined by the interaction of the fundamental policy objectives and the basic resource availabilities. If a particular resource is very scarce (that is, many alternative uses are competing for that resource), then its shadow price, or opportunity cost (the forgone benefit in the best available alternative that must be sacrificed), will tend to be high. If the supply of this resource were greater, however, the demand arising from the next best uses could be satisfied in decreasing order of importance, and its opportunity cost (or shadow price) would fall. Market prices will often reflect this scarcity correctly, but there is good reason to believe that in less developed countries imperfect markets may cause a divergence between market and shadow prices. Such divergences are thought to be particularly severe in the markets for three important resources: labor, capital, and foreign exchange.

Resource availabilities, however, need not be the only constraints operating in the economy: political and social constraints may be equally binding. The alternatives open to the government in pursuing its development objectives can be limited by these non-economic constraints to a narrower range than that implied by the basic resource availabilities. If the tools of general economic policy—that is, fiscal and monetary policy—cannot break these constraints, project analysis should take account of them by means of appropriate adjustments in shadow prices. For example, if the government is unable to secure a desired redistribution of income through taxation, it can use the allocation of investment resources as an alternative method of redistributing income. If in project analysis higher values were to be attached to increases in income accruing to the poorer groups within society, investment would be biased in favor of these groups. In other words, all available policy tools should be working jointly toward the same goals. If one instrument is inoperative or blunted, other instruments may be used to achieve the same end.

Project analysis is designed to permit project-by-project decisionmaking on the appropriate choices between competing uses of resources, with costs and benefits being defined and valued, in principle, so as to measure their impact on the development objectives of the country. In many cases, however, a more direct link is necessary with the sector and economy as a whole: for example, the merit of a project characterized by economies of scale cannot be judged without making an estimate of the demand for its output, and this in turn requires placing the project in its sectoral and country context.

Furthermore, in practice, many shadow prices (for land and natural resources, for example) are difficult to determine independent of the project appraisal process, because they depend on the alternative projects that have been rejected. This is the basic reason why a systematic scrutiny of plausible alternatives is at the heart of the appraisal process: it is not sufficient in practice to select an acceptable project whose benefits appear to exceed costs; it is necessary to search for alternatives with a larger surplus of benefits over identified costs. If such projects are found, it means that the opportunity cost of using, say, land in the project originally considered acceptable has been underestimated or wholly neglected.

Consideration of alternatives is the single most important feature of proper project analysis throughout the project cycle, from the development plan for the particular sector through identification to appraisal. Many of the more important choices are made at early stages when decisions are made concerning the alternatives that are to be rejected or retained for further, more detailed study. If economic analysis is to make a maximum contribution to the attempt to ensure that scarce resources are used to best ad-
vantage for the country, it should be used from the earliest phases of this process of successive sifting and narrowing down of options that are open to the country. The use of shadow prices reflecting basic policy objectives and resource constraints only in the final stage of appraisal, when most of the essential choices with respect to types of project and project design have already been made, tends to be mainly cosmetic. To be an effective aid in decisionmaking, shadow prices should also be used in framing sector strategies and in identifying promising project possibilities and designing their major features.
Further Reading

The literature on price theory and its applications is enormous, but some outstanding additional readings might be indicated within the most relevant categories of (1) microeconomics textbooks, (2) welfare economics, (3) cost-benefit analysis, and (4) economic analysis of projects.

1. Microeconomics Textbooks

The following "American style" textbooks may provide some background and also a fuller treatment for the topics discussed in chapters 1-3.

*Elementary Textbooks*


*More Advanced Textbooks*


2. Welfare Economics


Economics, vol. 72, no. 3 (August 1958), pp. 351–79.


3. Cost-Benefit Analysis


4. Project Appraisal


Index

Ability-to-pay concept, 214-16
Adelman, Irma, 236
Advertising, 16
Agricultural sector: developing economies and, 44, 45; development and, 3; distortions of incentives for 73-76; externalities and, 160; fragmented economy and, 139; interest rates on loans in, 90-95; liability for damage example and, 155-58; mechanization (Pakistan) and, 104-10; price policy and, 73, 74-76, 84-89, 100-04; price responsiveness in, 84-89; roads and location decisions in, 195; surpluses and, 30
Air pollution, 81
Average cost (AC): cost recovery and, 216-17; monopoly and, 143; pricing and, 175; short period efficiency and, 121-22; water supply pricing and long-run, 209
Average incremental cost (ACI): 205-06, 207; water supply pricing and, 208
Backward linkages, 161-62
Barone, Enrico, 50, 66
Bauer, Peter T., 47
Black market, 101
Bridge example of marginal cost pricing, 176-79
Budget: deficit in road, 195-96; public services and national, 111-12
Capacity, 171; electricity pricing and, 183-84, 187; marginal cost pricing and, 182; port pricing and, 200; pricing of urban services and excess, 78-79; production, 38; water supply pricing and, 210
Capital: development and, 46, 48-49; governments and price of, 234; opportunity cost of, 246, 247; reversing flight of, 95; shadow price of, 246; water supply pricing and indivisibility of, 208-09
Capital goods, 49, 53, 92; subsidies for, 73
Capitalists: ownership and, 68; price system and, 43; production functions and, 49
Capital market, liberalization and, 139
Choice, 209; class interest and development, 97; fundamental model of, 54-64; marginal cost pricing and, 179-80; policy instruments and, 236-40; price analysis and, 2; price system and, 43; social, 233; socialist economy and, 49, 51, 52; technological, 5-13; transport and factor, 98-100
Cigarette currency (P.O.W. price example), 39-42
Clark, J. M., 36
Cole, G. D. H., 53
Collateral, 94
Commodity agreements, 148-54
Commodity goods, 66, 139
Comparative advantage concept, 124, 126-32
Competition: in agricultural sector, 100; characteristics of "perfect," 114, 116; commodity agreements and, 148-54; comparative advantage concept and, 130-31; economic inefficiency and imperfect, 145; economic welfare and, 136-37; general equilibrium and "perfect," 71; imperfections of, 16; liability for damage and, 156; long-run equilibrium and "perfect," 123-24; marginal cost pricing and structure of, 201-02; resource allocation and, 169; restricted, 140-45; road user charges and, 196-97; short period efficiency and, 114-22; socialist economy and, 52, 53; supply and demand and, 23; water supply pricing and, 209-10
Competitive price system, 13-16
Congestion pricing, ports and, 82-83; roads and, 80-82
Consumer goods, 51
Consumer income distribution, 102-03
Consumers: intentional deficits and, 221; monopolist and, 144-45
Consumer surplus, 34-35
Consumers' valuation, 33
Consumption: collective, 164-65, 166, 167, 168; cost under a tariff, 135; electricity and new, 180, 228-30; externalities and, 149; food-grain prices and, 103; marginal rates of substitution in, 124; production-possibility frontier and, 7-8; public services and distribution of, 111, 113; socialist economy and, 51; water supply pricing and, 207, 211-12, 213
Consumption goods, 13-14
Correct price concept ("get the prices right"), 1, 235-36, 249
Cost-benefit analysis: in developing countries, 242-47; project analysis and, 252-53
Cost recovery, marginal cost pricing (transport example) and, 216-22
Costs: comparative, 126, 128, 132; law of increasing, 11-12; policy and, 231; of production, 35-38; road use charges and, 192-94; socialist economy and, 53; sunk, 207, 209. See also Average cost (AC); Average incremental cost (ACI); Long-run marginal cost (LRMC); Short-run marginal cost (SRMC); Textbook long-run incremental cost (TLRMC); Textbook marginal cost (TMC); Total cost (TC)
Credit, 41; agricultural sector and, 92-94, 102; for mechanization, 106
Currency (P.O.W. camp price example), 39-42
Damage, pricing system and, 155-58
Decisionmaking: choice and, 54-64; informational function of prices and, 65; investment, 199; roads and locational, 195; standard versus fundamental, 69-71
Default, agricultural loans and, 93
Deficits: electricity pricing and, 190; marginal cost pricing and, 181, 218-21; port pricing and, 200, 201; road user charges and budget, 195-96
Deflation, economic organization example of, 41
Demand: application of, 29-33; competition and, 23; competitive price system and, 14; competitive restrictions and, 141; cost recovery and, 217, 218; curve, 17; economic welfare implications and, 33-35; elasticity
Demand (continued)
of, 23–26, 27–29; electricity pricing and, 183–84, 187; energy pricing and, 191; equilibrium and, 19–23, 27–28; law of downward-sloping, 17–18, 22, 25; marginal cost pricing and, 176–77, 181, 182; monopoly and, 143–45; price instability and, 150, 151; pricing of urban services and excess, 79; production capacity and, 38; road use and, 194; schedule, 21–22; shifts in, 20; short-run marginal costs and, 198, 199; supply and prices interaction with, 7; water supply pricing and, 213

Developing countries: agricultural incentives in, 73–76; agricultural policy and, 100–02; commodity agreements and, 148–54; congestion pricing and, 82–83; cost-benefit analysis in, 242–47; economic responsiveness in, 44–45; externalities and, 159–64; interest rate policies in, 90–95; marginal cost pricing and, 170, 171; market information and, 72; oil prices and, 192; price policy and, 231–36; pricing of public urban services in, 77–80; project selection and, 247–49; subsidies and, 226; transfer pricing and, 146–48; transport technology and, 98–100; water supply pricing and, 215. See also Development

Development: agricultural, 3; class interest and, 97; market mechanism and, 45–49; pricing policy and, 1; programs for, 240; project appraisal and, 2; second-best policy and, 49, 149; strategies for rural, 73. See also Developing countries

Diminishing returns, law of, 8–11

Distributional effects of government expenditure (Colombia), 110–13

Distributional function of prices, 97–113

Distributive justice: economic efficiency and, 77; electricity pricing and, 190

Economic control systems, 67–68, 69, 71

Economic organization, P.O.W. camp example and simple, 39–42

Economic rent, 183

Economic welfare, 2; competition and, 136–37; harmful effects and, 157–58; project analysis theory and, 247; supply and demand and, 33–35

Economies of scale, mass production and, 11, 16

The Economics of Welfare (Pigou), 154

Education: as public good, 165, 167; public services and, 113; subsidies for, 112; water supply value and, 215

Efficiency: of agricultural resources, 74–75; competition model and, 136–37; competitive price system and, 15, 233; competitive restrictions and, 140; distributive justice and economic, 77; free trade and, 135; international commodity markets and, 148–49; long-run, 122–23; marginal cost pricing and, 173, 175–76, 207; market imperfections and, 137–38; port use and, 202; production-possibility frontier and, 12–13; resource allocation and, 170, 177; short period, 114–22; social justice and economic, 47; technological choice and, 6; water supply pricing and, 207, 209

Electricity: marginal cost pricing and, 180–81, 182, 183–84; marginal cost pricing and tariffs for, 184–90; subsidies for, 112; willingness-to-pay concept and, 227–28

Employment: agricultural prices and, 102–04; mechanization and, 107–10; in transport, 98–100

Environmental quality: congestion prices and, 81; infrastructural projects and, 163; liability for damage and, 154, 158; public goods and, 166, 168–69

Equilibrium: competitive price system and, 14; international trade, 127, 129; investment and general, 249; long-run, 123; market and general, 71–72; maximum profit, 120–21; monopoly and long-run, 143; perfectly competitive, 233; prices and, 20, 22, 23, 28, 33; socialist economy and economic, 51, 53; supply and demand and, 19–23, 27–28

Equity, 234; electricity rates and, 190, 228–29; road use charges and, 195

Exports, 124, 135, 138, 239; agricultural, 44, 100; comparative advantage concept and, 130–31; economic development and, 48; externalities and, 161; shadow prices and, 245

Externalities: in consumption, 149; cost recovery and, 217; defining, 153; developing countries and, 159–64; inputs and results without a price and, 69–70; public goods and, 165, 166, 167, 168, 169; public pricing policy and, 173; social cost and, 153–58; water supply pricing and, 214

Factors of production, 242; comparative advantage concepts and, 130, 132; costs of production and, 38; harmful effects analysis and, 158; law of diminishing returns and, 10

Fertilizer subsidy analysis, 222–27

Financial liberalization, 95–97

Financial resources, electricity rates and, 190

Fixed costs (FC): cost recovery and, 216–17; defining, 118; short period efficiency and, 118–20, 121

Foreign exchange, 73, 95, 96; labor-intensive techniques and, 98; less developed countries and, 2, 3; quantitative restrictions and, 238; shadow price of, 243

Forward linkages, 160, 162

Free trade, 132–36

Galbraith, J. K., 16

Government: economic development and, 44, 47, 48; expenditure on public services (Colombia), 110–13; financial liberalization and, 95–97; fragmented economy and intervention of, 138–39; getting prices right and, 235–36; income distribution and, 178–79; investment, 239

Growth, 46, 48, 234, 250, 252

Harmful effects, pricing system and, 155–58

Hayek, Friedrich von, 50

Health, subsidies for, 112

Housing, public, 165, 168

Imports, 125, 239, 240; comparative advantage concept and, 129–30; economic development and, 48; energy-efficient equipment, 191; equipment, 99; externalities and, 161; farm machinery, 106, 107; monopoly rents and, 96, 97; protection and, 132, 135, 139; shadow prices and, 245; wheat, 101

Incentives: distortions of agricultural, 73–76; economic growth, 46; investment, 239; private sector, 232; wages and, 235
Income, 44; defining peasant, 88; family and labor, 235; food prices and, 100; redistribution of, 253; returns to farming and, 75-76; socialist economy and, 53; water supply payment and, 215; willingness-to-pay and, 228-30

Income distribution: agricultural price policies and, 73; 100-02, 102-04; competitive price system and, 15-16; development and, 46, 47-48; economic efficiency and, 233; farm, 102; legislative optimal, 248; marginal cost pricing and, 178-79; prices and, 97; profitability analysis and, 242; project appraisal and, 250, 251-52, 253; redistribution and, 168-69; socialist economy and, 51; subsidies and, 221; urban services and, 78

Income transfers, 93

Increasing (relative) costs, law of, 11-12

Industrialization, 96, 232; externalities and, 159-60; forced, 47; overemphasizing, 159

Industry, 69, 73; development and, 3; externalities and, 162-63; "perfect" competition and, 114; prices and nationalized, 178, 179; roads and location decisions in, 195

Inefficiency, 6-7, 136, 177; analysis of, 145-46

Infant industry argument, 159

Inflation, 96, 242; agricultural loans and, 90; economic organization (P.O.W. camp example) and, 41

Information: as collective good, 167; cost of, 173; economic market and, 47; lack of accurate agricultural (Peru). 101-02; marginal cost pricing and, 175; market signaling and inadequate, 138; prices and, 65-72; shortages and surpluses and, 149

Infrastructure, externalities and, 160, 163

Input: agricultural subsidies for, 102, 222, 223, 225; externalities and, 160-62; production, 8-11

Interest rates, 48, 89, 95, 237; on agricultural loans, 90-95; mechanization credit and, 106

Internal rate of return criterion (project analysis), 246-47, 248

International trade, 48-49, 124-25, 127, 129

Investment, 199; average incremental cost and, 205; general equilibrium and, 249; the government and, 48; government versus private, 239; marginal cost pricing and, 203-04; mechanization and, 105, 106; national, 250; new technology and agricultural, 75-76; planning and, 244; policy instruments and, 238, 240; private agricultural (India), 101; public, 232; in roads, 112, 113, 194-95, 196; textbook long-run incremental cost and, 204-05; "urban bias" of, 113; water supply pricing and, 210

Investment planning, 160

Invisible hand doctrine, 136-37

Johnson, D. Gale, 74

Knight, Frank H., 158

Knowledge, 47; economic analysis and limited, 65-66; pricing and imperfect, 202; water supply and consumer, 214

Labor, 234; development and, 48-49; mechanization and, 105, 107-10; shadow price of, 245-46

Labor-intensive techniques (transport), 99-100

Laissez-faire, 158, 237; perfect competition and, 136-37

Lange, Oskar, 66

Least-cost solution, 179

Lerner, Abba, 66

Licensure: import, 97; rice-milling (India), 101; road-pricing scheme and, 80-82

Loans, agricultural, 90-95

Locational decisions, 195

Long-run marginal cost (LRMC): bogus nature of, 177; cost recovery and, 217, 219; efficiency and, 122-23; electricity pricing and, 188; port pricing and, 200; pricing and, 187; road user charges and, 194, 197; water supply pricing and, 208, 209, 210. See also Short-run marginal cost (SRMC)

Managers, socialist economy and, 51-53

Manufacturing, 73

Marginal capacity cost (MCC), 204, 205; water supply pricing and, 208

Marginal cost: application of theory of, 179-82, 217-18; bridge example and practice of, 176-79; comparative advantage concept and, 127; competitive restrictions and, 141-44; competitive structure and, 201-02; cost recovery example (transport) and, 217-22; deficits and, 218-21; defining, 115-16, 171, 173, 181, 203-07; departures from, 227-30; electricity and, 180-81, 182, 183-84; electricity tariffs and, 184-90; energy and, 180-81, 190-92; pricing urban services and, 78-79; principles of public pricing and, 171-76, 216-17; ports and, 199-203; road use and, 192-99; short period efficiency and, 114-22; subsidy example (fertilizer) and, 222-27; textbook, 203-04; water supply pricing and, 207-16. See also Long-run marginal cost (LRMC); Short-run marginal cost (SRMC)

Marginal rate of substitution (MRS), 62, 63, 123

Marginal rate of transformation (MRT), 62, 63, 123, 124; resource allocation and, 233

Marginal revenue (MR); competitive restrictions and, 141-44; profit maximizing equation and, 122-23

Marginal social cost (MSC), 247-48, 249

Marginal social value (MSV), 247-48, 249

Market failure, 241; government intervention and, 138-39; public goods and, 167-68; public sector and, 169; underdevelopment and, 137-38

Marketing controls (agriculture), 101-02

Market mechanism, development and, 45-49

Markets: commodity agreements and commodity, 148-54; economic organization of simple (P.O.W. camp example), 39-42; information flow and planning and, 66-72; interdependence of, 2; limited knowledge and, 65-66; new electrical, 228-30

Marshall, Alfred, 27-28

Marx, Karl, 11, 71

Mechanization, 99; consequences of tractor (Pakistan), 104-10; fragmented economy and, 139

Metering, 168, 188-89, 210-11

Migration, 246, 249

Mises, Ludwig von, 49-50, 53, 66

Mobilization function of prices, 83-97

Monopoly: competitive price system and, 16, 137; in developing economies, 44, 45; economic organization (P.O.W. camp example) and, 40, 42; efficient alloca-
Monopoly (continued)
ion and, 209; international markets and, 149; public goods and, 168; restricted competition and, 140-45; road user charges and, 136
Monopoly rents (imports), 96, 97
Nationalization, 178, 179
Net present value criterion (project analysis), 246-47, 248
Oligopoly, 140, 141, 145; international, 147, 148; price stabilization and, 152
Opportunity costs, 158, 171, 252; of capital, 246, 247; comparative advantage concept and, 127; defining, 36, 37; public services and, 112; shadow price derivation and, 245; supply and demand and, 33; transfer pricing and, 147
Outlay costs, 36-37
Output: electricity, 183; externalities and, 159-60; farm, 75; marginal cost pricing and, 203; money sector, 44; price responsiveness and agricultural, 86-89; production; 8-11; public goods and, 169; socialist economy and, 51-53; supply and demand analysis and, 33
Ownership: costs of, 38; in mixed economies, 68-69; private, 50; public, 49, 52
Pareto optimum, 13, 123-24, 135, 169, 233, 247, 248
Pareto, Vilfredo, 66
Patent laws, 139
Planning: central, 46, 50, 52, 66, 67, 68-69, 71-72; cost-benefit analysis and national, 243-44; demand forecast (electricity) and, 157; development and market mechanism and, 45-46, 48; investment and, 160; market economy and informational flows and, 66-71; port pricing and, 201; price policies and, 232; price system and central, 43; short-run average cost and, 122
Policy: correct prices and development, 1, 235-36, 249; costs and, 231; developing economies and, 44, 47, 48; externalities and, 154; financial liberalization, 95-96; financial repression, 96-97; instruments for, 236-41; interest rate, 90-95; public goods and, 164; water supply pricing and, 207. See also Price policy
Political relations, market and planning comparisons and, 68-69
Ports: congestion levies and, 82-83; marginal cost pricing and, 199-203
Possibility frontier. See Production-possibility frontier
Power relations, market and planning comparison and, 68-69
Preferences: choice and, 62, 63, 64; decisionmaker's, 57-60; price system and, 43; socialist economy and, 51
Premiums, payment of, 45
Present work of incremental system cost (PWISC), 205, 206
Price fixing: by Central Planning Board, 53; by law, 29-30
Price policy: agricultural, 73, 74-76, 84-89, 100-04; application of marginal cost pricing and, 179-82; cost recovery and transport, 217-22; developing economies and, 44, 231-34; electricity and, 183-90; energy and, 190-92; fertilizer subsidy and, 223-25; marginal cost pricing in practice and, 176-79; ports and, 199-203; public pricing principles and, 171-76; road use and, 192-99; theories of, 234-36; urban services and, 79-80; water supply and, 207-16. See also Policy
Prices: absolute, 44; agricultural incentives and, 74-76; allocation and, 72-76; competitive, 136-37; competitive system of, 13-16; congestion, 80-83; correct, 1, 235-36, 249; differential, 35; distributional effects and, 97-113; downward-sloping demand and, 17-18; economic organization (P.O.W. camp example of), 39-42; economics and, 23; elasticity of supply and, 27; equilibrium and, 20, 22, 23, 28, 35, 33, fixed, 29-30, 53; free-market, 74; information and, 65-72; mobilization and, 83-97; monopoly and, 16; multiple-correlation analysis and changes in, 74; public urban services and, 77-80; rationing and, 76-83; relative, 44, 45; relative agricultural, 102-04; responsiveness to, 1-2; socialist economy and, 49-54; a specific tax and, 30-32; supply and demand interaction with, 71; transfer, 146-48; uniform, 33-34, 35. See also Marginal cost pricing; Price policy; Price system; Shadowprices
Price stabilization, 149-52
Price system: characteristics of competitive, 13-16; economic functions of, 46; liability for damage and, 155-57; nature and necessity of, 42-44; social significance of, 233-34
Prisoner-of-war (P.O.W.) camp price example, 39-42
Private sector: agricultural investment and, 75-76, 101; fragmented economy and, 138; investment and, 239; price policy and, 232; road costs and, 193
Production: agricultural, 73, 74-76, 102, 104; central and market economies comparison and, 69; costs of, 35-38; decisionmaking and, 65; development and, 46; diminishing returns and, 8-11; economies of scale and mass, 11, 16; equilibrium system and, 14; free trade and, 134, 135; low-cost food and control of, 100-01; marginal cost pricing and, 171-72; policy instruments and, 238, 239; price system and, 43; restricted competition and, 140; socialist economy and, 49, 51-53; supply and costs of, 35
Production function shifts, agricultural incentives and, 75-76
Production-possibility frontier, 233; choice and, 56-57, 61-62, 63, 64; comparative advantage concept and, 127-28; competitive price system and, 15; competitive pricing and, 137; diminishing returns and, 8-11; efficiency and, 12-13; increasing (relative) costs and, 11-12; technical efficiency and, 124; technological choice and, 6-8
Profit, 45; economic cost and normal, 38; income distribution and, 179; long-run efficiency and, 122-23; marginal cost pricing and, 178, 179, 181; maximization of, 149; monopoly and, 141, 142, 146; new agricultural techniques and, 73; price setting and, 16; project analysis and, 253; public utilities and, 209 n31; short period efficiency and, 114, 116-17, 120-21, 122
Profitability analysis, 242
Project appraisal, 2; calculation technique and, 246-47; choice model and, 64; cost-benefit analysis in, 241-47; objections to methodology for, 251-54; preliminary theory of, 247-49; problem of, 177; traditional methods of, 250-51
Projects: electrification, 186, 187, 229; externalities and, 159–64; interconnected, 240; selection of, 247–49; water supply, 207, 208
Protection, 132–36; import tariffs and, 139; as policy instrument, 238–39
Public goods: classification of, 166–69; defining, 164–65; inputs and results without a price and, 68–70; nature of, 164–66; pricing and, 172
Public housing, 165, 168
Public pricing. See Marginal cost pricing
Public sector: agricultural investment and, 76; market failure and, 169; pricing principles and, 171–76; public services and, 77–80, 110–13
Quantitative measures, 236–40
Quotas, 238
Railways, 180; road pricing and, 196–97
Rates (service tariffs): deficits and, 221; electricity, 180, 182, 184–86, 188, 189, 190, 191; marginal cost pricing and, 174–75; ports and, 202; water supply, 212–13; willingness-to-pay concept and, 227, 228
Rationing, 30; credit, 94–95; electricity pricing and, 187; of foreign exchange, 96; prices and, 76–83, 175; suction and planning and, 69; the very poor and, 101; water supply, 213
Research and development (pharmaceuticals), 147
Resource allocation: agricultural loans and, 92; competition and, 169, 201–02; defining optimal, 114; electricity tariffs and, 185; international economy and optimal, 124; marginal cost pricing and, 170, 171–72, 177, 209; marginal equalities and, 63; marginal rate of transformation and, 233; perfectly competitive economy and, 123; price changes and, 231; prices and, 72–76; project analysis and, 249, 252, 253; public goods and, 167; socialist economy and, 49, 50–53; sunk costs and, 207, 209; technological choice and, 5–6; underdevelopment and, 137–38; water supply pricing and, 207, 210
Resources: agricultural pricing policy and efficient use of, 74–75; utility service and saving of, 229
Ricardo, David, 124, 126–32, 135
Risk, 149; decision problem and, 69; innovation and, 147
Roads, 100, 180; competition and, 202; costs of using, 192–97; externalities and, 160, 163; investment in, 112, 113, 194–95, 196; marginal cost pricing and, 192–99; mechanism and, 99; rationing function of prices and, 80–82
Robbins, Lionel, 50
Rural sector: fragmented economy and, 139; investment “urban bias” and, 113; water supply pricing example and, 210–14
Savings, 89, 92, 139; financial liberalization and, 95; government and, 48; less developed countries and, 2–3; returns to farming and, 75–76
Second-best policy, 235; development and, 49, 149; electricity rates and, 189; shadow prices and, 248; water supply prices and, 209–10
Services: pricing of public (Colombia), 110–13; pricing of public urban, 77–80; productive, 14; public goods arising from nonmarketable, 166–68
Service tariffs. See Rates
Shadow prices: choosing, 249; derivation of, 244–46; electricity prices and, 189–90; of foreign exchange, 243; for labor, 99; project appraisal and, 250–54; project selection and, 248; water supply and, 209–10
Shortages, 53, 149; controls and, 238; market, 76; price ceilings and, 29–30
Short-run marginal cost (SRMC): bogus nature of, 177; defining, 198–99; marginal cost pricing definition and, 173, 204; port pricing and, 200–01; pricing and, 187; profit maximization equation and, 122; road user charges and, 194, 197; short period efficiency and, 114–22; water supply pricing and, 208. See also Long-run marginal cost (LRMC)
Singapore, congestion pricing example and, 80–82
Smith, Adam, 1, 124–25, 136–37
Social benefits: mechanism and, 107; public pricing policy and, 172, 173, 174
Social costs, 171; externalities and, 159–64; marginal, 247–48; mechanization and, 107; road pricing and, 80, 193
Socialist economy: compared with market economy, 66–72; ownership and, 68–69; prices in, 49–53, 66
Social welfare, 148
Spatial complexes, externalities and, 162–63
Standard of living, 69
Stigler, George J., 135
Strict marginal cost (SMC), pricing and, 208
Subsidies, 94, 139, 230, 238, 248; agricultural, 45, 102; capital goods, 73; credit, 92–93; deficits and, 220; fertilizer example of, 222–27; fuel, 192; imported wheat, 101; protection and, 134–35; public housing, 168; public service, 110–11, 112, 113; reasons for, 221–22; supply and demand analysis and, 32
Substitution, 2; downward-sloping demand and, 18; farm labor, 109–10; foodstuffs and, 100; marginal rate of, 62, 63, 123–24; price setting and, 16; production-possibility frontier and, 6
Supply: application of, 29–33; competition and, 23; competitive price system and, 14; costs of production and, 35; demand and prices interaction with, 71; economic welfare implications and, 33–35; effect of agricultural price changes on, 84–89; elasticity of, 26–27; equilibrium of, 19–23, 27–28; price instability and, 150–52; schedule, 18–19; shifts in, 20; short period efficiency and, 116–17; short-run marginal cost and, 198; upward-sloping curve of, 19
Surpluses, 53, 149; agricultural, 88; consumer, 34–35; price floors and, 30; producer, 35
Tariffs: electricity, 184–90; import (Latin America), 139; as policy instrument, 238–39; protection and, 132, 133, 135. See also Rates (service tariffs)
Taxes, 139, 237, 238, 239, 248; ad valorem, 32–33; on commodities, 76–77; on electricity, 190; financing of losses and, 177, 179; on fuels, 189, 191; supply and demand analysis and, 30–32; utility tariffs and, 228
Taylor, Fred M., 50
Technological change: agricultural incentives and, 75; transport and, 98–100
Technology, 102; agricultural, 74, 75; choice of 5–13, farm machine, 104–10; imports and, 48; loans and agricultural, 92, 93; monopoly and, 142–43; Pharmaceuticals and innovation in, 147, 148; price system and, 43; vertical linkages of, 163
Tenants, mechanization and, 107, 108, 109
Textbook long-run incremental cost (TLRIC), 203, 204, 206
Textbook marginal cost (TMC), 203–04, 206
Time lags, market signals and, 167–68
Total cost (TC): defining, 119; long-run average, 123; short period efficiency analysis and, 118–20, 121
Total dollar revenue charges, demand elasticity and, 24–25
Tractors. See Mechanization
Trade: free, 132–36; international, 48–49; 124–25, 127, 129
Training: market imperfections and, 138, 139; public services and, 113; technical skills and, 159
Transaction costs, 168
Transnational corporations, transfer pricing and, 145–48
Transportation: competition and, 202; congestion pricing and, 80–82; cost recovery and, 217–22; externalities and, 163; labor-intensive techniques and, 98–100; marginal cost pricing and, 180, 192–99; price responsiveness and agricultural, 89; public, 80, 81, 82
Uncertainty, 67, 149; decision problem and, 69; private investment and, 239
Unemployment, 92, 235, 239; financial liberalization and, 95; foreign exchange and, 98; fragmented economy and, 139; mechanization and, 109; production-possibility frontier and, 6–7
United Nations Conference on Trade and Development (UNCTAD), price stabilization and, 149–52
Urban sector, pricing of public services in, 77–80
User charges, 168; redistribution of income argument and, 221; road, 192, 193–98; short-run marginal cost and, 199
Variable costs (VC): cost recovery and, 216–17, 218, 219–20; short period efficiency and, 119–120, 121
Wages, 235, 249
Water supply: marginal cost pricing and, 207–16; pricing of, 77–79
Wicksteed, P. H., 49
Willingness-to-pay concept: analysis of, 227–30; pricing of public urban services and, 77; water supply pricing and, 214–15

The full range of World Bank publications is described in the Catalog of World Bank Publications; the continuing research program is outlined in World Bank Research Program: Abstracts of Current Studies. The most recent edition of each is available from the Publications Distribution Unit, The World Bank (Dept. B), 1818 H Street, N.W., Washington, D.C. 20433, U.S.A.

Gerald M. Meier is professor of international economics at Stanford University and a consultant to The World Bank.
This book is for the development practitioner who wants to learn more about the role of prices in a country's development process. It explains the essential elements of a price system, the functions of prices, the various policies that a government might pursue in cases of market failure, and the principles of public pricing of goods and services provided by government enterprises. Pricing Policy for Development Management provides project evaluators with an appreciation of the underlying logical structure of cost-benefit project appraisal.

To integrate the theoretical, applied, and policy dimensions of price analysis, the editor has combined his commentary with selected readings that elucidate the basic principles of pricing and illustrate their application to development problems. Many of the readings are drawn from the experience of development practitioners in the fields of agriculture, industry, power, urban services, foreign trade, and employment. The principles outlined are therefore relevant to a host of development problems.

Although the whole of this book is greater than the sum of its parts, the reader may select portions to study according to individual needs and interests.