DISCUSSION PAPER

NEOCALSSICAL ECONOMETRICS: NON-NEGATIVITY CONSTRAINTS

by

Michael J. Hartley

October 1985

Development Research Department
Economics and Research Staff
World Bank

The World Bank does not accept responsibility for the views expressed herein which are those of the author(s) and should not be attributed to the World Bank or to its affiliated organizations. The findings, interpretations, and conclusions are the results of research supported by the Bank; they do not necessarily represent official policy of the Bank. The designations employed, the presentation of material, and any maps used in this document are solely for the convenience of the reader and do not imply the expression of any opinion whatsoever on the part of the World Bank or its affiliates concerning the legal status of any country, territory, city, area, or of its authorities, or concerning the delimitations of its boundaries, or national affiliation.
NEOCLASSICAL ECONOMETRICS: NON-NEGATIVITY CONSTRAINTS

by

Michael J. Hartley*

Development Research Department
World Bank**

October, 1985

* This paper is based on portions of Hartley (1981a, 1981b, 1983a). The author is indebted to Arne Drud for helpful discussions.

** The World Bank does not accept responsibility for the author's views. These should not be attributed to the World Bank or its affiliated organizations. The findings, interpretations and conclusions are the results of research supported by the Bank; but do not necessarily represent its official policy.
Abstract

The paper considers the problem of determining the parameters in equality-constrained neoclassical economic models in which the decision variables are also constrained to be non-negative. Such problems frequently arise in models of household and farmer behavior. The household maximizes utility subject to income and/or time constraints and the fact that its demand for various goods and services and its supply to various segments of the labor market are non-negative. The profit maximizing farmer allocates a given amount of land over a set of possible crops, where the land allocation, input demand and output supply associated with each crop are non-negative. The paper formulates a canonical form of this model and discusses the nature of the inverse (neoclassical econometric) problem. The traditional (regression based) econometric approach is generalized to a multi-decision situation and the computational difficulties and inherent paradigmatic limitations are discussed. A simple alternative deterministic neoclassical econometric approach, avoiding such problems, is proposed and a simple algorithm is discussed. Finally, extensions of the model and topics for future research are considered.
# Table of Contents

I. Introduction ................................................... 1  

II. The Standard Neoclassical Model with Non-Negativity Constraints  
    A. The Maintained Hypothesis (Rationality) ..................... 3  
    B. Two Simple Illustrations ..................................... 4  
        1. The Representative Household ............................. 4  
        2. The Competitive Farm ...................................... 5  
    C. Nonlinear/Dynamic Programming ................................ 7  
    D. The Inverse Problem ......................................... 8  

III. The Traditional (Regression-Based) Approach .................... 9  
    A. General Considerations ......................................... 9  
    B. The Representative Household ................................ 12  
        1. The Two-Good Case ........................................ 12  
        2. The Three-Good Case ..................................... 15  
    C. Generalization ................................................ 19  

IV. A Neoclassical Econometric Approach .............................. 20  
    A. A Deterministic Model ........................................ 21  
    B. A Simple Algorithm ........................................... 22  
    C. Discussion .................................................. 25  

V. Conclusions, Extensions and Future Research ...................... 27  
    A. Conclusions ................................................ 27  
    B. Extensions and Future Research .............................. 29  

---

1/ I would like to thank Miriam Bailey for her excellent word-processing.
I. Introduction

Neoclassical micro-economic theory is based upon the maintained hypothesis that individual decision-making units -- households, firms, etc. -- maximize a specified objective function or goal, g, with respect to a decision or policy vector, d, given the value of a state vector, x, and, in a general equilibrium model, subject to the fact that their decisions are not independent of those of their neighbors. In many such applications, the decision vectors are, by definition, non-negative. For example, in household expenditure surveys with even moderate disaggregation, the investigator will typically encounter numerous instances of individual units spending no income on one or more categories of goods and services over the duration of the study. Similar situations obtain, for example, in the allocation of land, input-demand and output-supply over all crops in surveys of agricultural producers.

We shall utilize these two examples in developing general econometric methods for such allocation problems involving non-negativity constraints on the domain of the decision-variables. Such problems have come to be termed multivariate Tobit-type models (following the seminal work of Tobin (1958)) or, more generally, Limited Dependent Variable (LDV) econometric models (see, e.g., Amemiya (1984) or Maddala (1983)). These models have their origins in the linear/nonlinear regression model, combined with the problem of estimating the parameters of a "censored" probability density function (p.d.f.) in statistical theory. Such methods will be termed the "traditional (regression-based) approach."

In section II we shall begin by formulating the maintained (neoclassical) hypothesis governing the behavior of the individual economic
unit. We then illustrate by exhibiting two standard examples -- the utility-maximizing household, subject to an income constraint; and the profit-maximizing farmer, subject to the constraints of a given amount of land and a given technology associated with each possible crop or land use. We then turn to a brief review of available nonlinear and dynamic programming algorithms to compute the optimal decision-values in the situation where the parameters are known. Finally, we formulate the nature of the inverse problem confronting the investigator.

In the next section, we consider application of the traditional (regression-based) approach to these problems. The essence of this stochastic model involves introduction of a vector of (normal) random errors, added to the optimal (unconstrained) decision functions. If infeasible, these must then be "mapped" back onto the boundary of the feasible decision space. We formulate the "closest boundary value" rule to define a unique mapping. Apart from the implausibility of the underlying paradigm, implementation of this approach quickly breaks down in problems of moderate dimension, as the order of multiple numerical integrations required to estimate the "true" parameters soon exceeds present computer capabilities. We illustrate these difficulties in terms of our leading example.

In section IV we introduce an alternative neoclassical econometric approach to this class of problems. We consider a deterministic model, and formulate the inverse problem. A simple algorithm is proposed to calibrate the unknown parameters.

Some concluding remarks and a discussion of possible extensions and future research are reserved for section V.
II. The Standard Neoclassical Model

A. The Maintained Hypothesis

We assume (here) that individual economic units have complete knowledge of the Sxl state vector, $\mathbf{z}$; the Pxl parameter vector, $\mathbf{\theta}$; the objective function or goal, $g$; and the Cxl equality-constraint vector function, $c$. Each unit then chooses a Dxl decision vector, $d$, to maximize the objective function,

$$g(d, z, \theta),$$

subject to the $C < D$ functionally-independent equality constraints,

$$c(d, z, \theta) = 0,$$

and the $D$ non-negativity constraints,

$$d \geq 0.$$

We assume that $g$ is quasi-convex in $d$ for all values in some region, $D^0$, including the region, $d'd < c^0$ for a suitably large constant $c^0 > 0$.
containing the non-negative orthant, \( R^0_0 \); 1/ that the vector, \( z \), belongs to the state space, \( S^0 \); and that the parameter vector, \( \theta \), is restricted to a compact parameter space, \( \Theta^0 \). Further we assume that the functions, \( g \) and \( \phi \), are twice continuously differentiable in \( z \) and \( \theta \) over \( D^0 \times S^0 \times \Theta^0 \). 2/

B. Two Simple Illustrations

1. The Representative Household:

We consider (here) only the income-allocation aspects of the "representative" household. 3/ Each economic unit maximizes the utility function,

\[
g(d, z^{(1)} , \theta) , \tag{2.4}
\]

subject to a budget constraint,

\[
(2)' \cdot d - z^{(3)} = 0 \, . \tag{2.5}
\]

1/ For example, if \( g \) denotes a utility function, and we wish to model cases where non-negativity constraints on \( d \) may be binding, then \( g \) must be defined over negative values of certain elements of \( d \). This would appear to rule out for present purposes such specifications as the direct "trans-log" utility function (Christenson, Jorgenson, Lau (1975)), involving only the logarithms of elements in \( d \) -- see section IV.

2/ Some of these classical assumptions, however, are not required in certain of the methods -- see below.

3/ Generalization of the model to incorporate, say, time-allocation, involving a non-negative labor force participation/labor supply decision by household members -- zero, if not; positive, if labor is supplied -- is straightforward, see Hartley (1985b).
and the restriction that all demands are non-negative,

\[ d > 0 , \]

where 

\[ g \]

is any "well-behaved" utility function,

\[ d = \text{demand vector for goods and services}, \]

\[ s^{(1)} = \text{household characteristics vector}, \]

\[ s^{(2)} = \text{price vector for goods and services (also } p), \]

\[ s^{(3)} = \text{income (also } y). \]

Thus, here we have the state vector,

\[ s = [s^{(1)}' s^{(2)}' s^{(3)}]' . \]

2. The Competitive Farm:

We assume the farmer operates a holding of a given cultivable area, on which the land must be allocated among \( K \) possible crops -- each of which has a known technology. Each unit maximizes profits,

\[ s = \sum_{k=1}^{K} \{ s^{(1)} k \cdot d^{(1)} k - s^{(2)} k \cdot d^{(2)} k - s^{(3)} k \cdot s^{(4)} k \} = \sum_{k=1}^{K} g_k , \]

subject to the \( K \) constraints of a given technology for each crop,

\[ d^{(1)} k - f_k (d^{(2)} k, d^{(3)} k, s^{(4)} k, s_k) = 0 , \]

\[ \]

\[ \]

\[ \]

\[ \]

\[ \]

\[ \]

\[ 1/ \text{ Alternative notation is introduced for subsequent purposes.} \]
with \( k = 1, \ldots, K \); the land-allocation restriction,

\[
\sum_{k=1}^{K} d_k(3) - s_k(6) = 0 ;
\]  

(2.10)

and the \( D \) overall non-negativity constraints,

\[
d = [d(1)', d(2)', d(3)']' \geq 0 ,
\]  

(2.11)

with \( d(1) \equiv \text{vec}(d_k(1)) \), \( s(1) \equiv \text{vec}(s_k) \), and where

- \( d_k(1) \) = output of crop \( k \)
- \( d_k(2) \) = variable input vector for crop \( k \)
- \( d_k(3) \) = land-allocation vector over \( K \) crops
- \( s_k(1) \) = output price for crop \( k \)
- \( s_k(2) \) = variable input price vector
- \( s_k(3) \) = fixed input price vector
- \( s_k(4) \) = fixed input vector for crop \( k \)
- \( s_k(5) \) = holding/farmer characteristics vector
- \( s_k(6) \) = area of holding

and \( d(2) \equiv \text{vec}(d_k(2)) \) "stacks" the \( d_k(2) \) in a vector. \(^1\)

\[^1\] A dynamic multi-period version of this model, applicable to mixed holdings involving both annuals and perennials and utilizing a Markovian land-rotation system, is given in Bellman and Hartley (1985).
C. Nonlinear/Dynamic Programming

In the event that an "interior solution" obtains to the canonical form of the problem under the maintained hypothesis, then classical methods may be employed to find the solution, \( d^* \). However, in many micro-economic applications, one or more of the inequality constraints will be binding; and, for a given value of \( \theta \) and given "well-behaved" specifications of the functions, \( g \) and \( c \), optimal feasible values for \( d^* \) can only be computed as the solution to a nonlinear and/or dynamic programming problem.

This is a long-standing problem -- see, e.g., Hadley (1964) for a review of the early literature. Various nonlinear programming algorithms may be employed, and no single method, at present, appears dominant in all applications. The reduced gradient method (Lasdon, Waren, Jain and Ratner (1978); the NAG Libraries software package, SALQDR; and the SOL/NPSOL Fortran package, involving solving a sequence of quadratic programming problems (Gill, Murray and Wright (1981); Gill, Murray, Saunders and Wright (1983), etc.) are widely recommended. See also Drud (1985).

Finally, as we subsequently note, many of the allocation-type problems in mathematical economics with non-negativity constraints can be formulated as dynamic programming problems (see, e.g., Bellman (1957, 1963)).

---

1/ See, e.g., Hartley (1985b) for a parallel discussion of the case where inequality constraints in (2.3) are removed or are unnecessary (Christenson, Jorgenson, Lau (1975)).

2/ Available from NAG Libraries, 1250 Grace Court, Downers Grove, Illinois, USA.
and Bellman and Dreyfus (1962). Such methods, in avoiding any use of the classical methods of the calculus, thus permit a richer array of potentially more realistic maintained hypotheses -- a point repeatedly stressed by Bellman. This matter assumes even greater significance as we generalize the nature of the domain constraints, since the state of computer technology continues to escalate by leaps and bounds.

D. The Inverse Problem

The problem confronting the investigator, however, is the inverse of that of the decision-maker under the maintained hypothesis. Suppose we are given a suitably-sized sample/set of data, on the actual decisions, \( d_{it} \), taken by unit \( i \) in period \( t \) (with \( N, T \geq 1 \)), when decision-makers are confronted with the "state of nature", \( s_{it} \). Then the problem for the investigator is to determine the unknown parameter values, \( \theta \), under particular specifications of \( g \) and \( \zeta \), which "best approximate"...
the observed data, \((d_{i\tau}, s_{i\tau})\). 1/

III. The Traditional (Regression-Based) Approach

A. General Considerations

In terms of its historical antecedents within the statistical literature, the problem of estimation of the parameters of a (non-negative) censored p.d.f. apparently was considered first by Pearson and Lee (1908), with later important contributions by Fisher (1931), Hald (1949), and Cohen (1949, 1957) -- generally in the context of a univariate normal distribution. In the econometric literature, Tobin's (1958) seminal paper, extended the analysis to a censored normal regression model (motivated by the problem of modeling the demand for a consumer durable good); and Amemiya's (1973) influential discussion of the asymptotic statistical properties in such (independent, but not identically distributed) cases, spawned considerable interest among economists in so-called LDV models, and their generalization to multi-decision economic problems. 2/

In terms of the applicability of LDV models to neoclassical economic models of the form, (2.1)-(2.3), certain desiderata should be applied. First,

1/ In this connection, recent research by Gallant and Golub (1984) and Diewert and Wales (1984) in developing classes of so-called "flexible functional forms" which impose "curvature restrictions" -- quasi-convexity, quasi-concavity, etc. -- on the functional form of the objective function (or its counterpart under duality) are most important contributions -- both for the programming problem in the "forward solution", as well as its inverse.

2/ See Amemiya (1984) and Madalla (1983) for comprehensive reviews of this literature.
note that in a multiple-decision context the deterministic form of the
maintained hypothesis suggests that the same parameters, $\theta$, in the utility,
production, etc., functions may appear in all decision functions -- whether
constrained by inequalities or not. Thus, in general, there are complicated
nonlinear cross-equation parameter restrictions on the system of decision-
functions. However, in the present (non-linear/dynamic programming) problem,
where analytic solutions, in general, are not feasible, it will not be
possible for the investigator to formulate a closed-form expression for the
functional form of the "regression equations", which exhibits the appropriate
cross-equation restrictions.\footnote{E.g., see Heckman (1974, 1976, 1979),
where no parametric restrictions are imposed on the participation and labor
supply equation due to formal optimization of a utility functions.}
Rather, "regression functions" are formulated in an ad hoc manner, and concern
has been focused upon assuring that the (binary/non-negative, etc.) domain of
the dependent variables is suitably accommodated.
Second, even if cross-equation restrictions were imposed, for
example, by ignoring the inequality constraints, (2.3), the underlying
paradigm for a neoclassical economic model is still somewhat contrived. We
shall characterize this approach as follows (Hartley (1983a), (1984b)):

(1) relative to the objective function, $g$, of (2.1), and subject only to
the equality constraints, (2.2), an "optimal" decision, $d_{it}^*$, is
generated, ignoring all inequality constraints on $d_{it}$;\footnote{See Hartley
(1985b).}

(2) a random error vector, $\xi_{it}(1)$, is then added to (D-C) of the
functionally independent elements in the sub-vector, say, $d_{it}(1)^*$, to
account for stochastic "errors-of-maximization", "errors-in-
variables", etc.;

(3) two possible cases may obtain:
(a) if the random vector, \( d_{it}^{(1)*} + \xi_{it} \), satisfies all inequality constraints, then \( d_{it}^{(1)} = d_{it}^{(1)*} + \xi_{it} \) is presumed to represent the observed value of the decision sub-vector; or
(b) if the random vector, \( d_{it}^{(1)*} + \xi_{it} \), violates one or more inequality constraints, then a suitable vector of constants, \( v_{it}^{(1)} \), is added to the above to produce the observed value,
\[
d_{it}^{(1)} = d_{it}^{(1)*} + \xi_{it} + v_{it}^{(1)}
\]
where \( v_{it}^{(1)} \) is chosen to yield the smallest distance from \( d_{it}^{(1)*} + \xi_{it} \) to the boundary of the feasible domain;

(4) finally, given \( d_{it}^{(1)} \), the remaining \( C \) decisions, \( d_{it}^{(2)} \), are obtained by solving the \( C \) equality constraints in (2.2).

We shall illustrate the closest-boundary-value principle in generalizing the single-equation traditional econometric approach to the multiple-decision, inequality-constrained-domain case. It should be stated at the outset, however, that this paradigm -- namely, how the economic decision-making unit is presumed to arrive at the observed value, \( d_{it} = \begin{bmatrix} d_{it}^{(1)}' & d_{it}^{(2)}' \end{bmatrix}' \), in our view, is somewhat implausible. Why should the decision-maker: "(a) begin by solving an (inequality) unconstrained problem, while presumed to be aware of such constraints; and then (b) proceed to "fix-up" the domain in infeasible cases, following the introduction of stochastic disturbances?"
Indeed, as we shall see, this paradigm becomes considerably more complicated to implement as the number of decision variables in \( d_{it}^{(1)} \) increases. We believe that the traditional (regression-based) approach, while quite appropriate for statistical models in biometry, physics, engineering, and
other natural and/or physical sciences is unfortunately inappropriate for behavioral sciences, involving data generated by an underlying neoclassical (constrained optimization) model -- except in the univariate special case, where the approach can be made equivalent to certain neoclassical econometric alternatives.

v. The Representative Household

We shall employ our simple model of the representative household to illustrate the above propositions. To exhibit the distinction between the single and multiple-decision cases, we begin with the 2-good problem; and then examine the 3-good case, as a precursor to discussion of the D-good analogue.

1. The Two-Good Case:

With D=2, we first solve (2.5) for \( d_2 \) and substitute into (2.4), yielding the income-constrained direct utility function,

\[
\phi(d_1, (y-p_1 \cdot d_1)/p_2, \varphi^{(1)}, \varphi^{(2)}) \tag{3.1}
\]

We then find the value, \( \hat{d}_1 \), which maximizes (3.1), and may be obtained as a solution to:

\[
\frac{3\varphi}{3d_1} - \frac{3\varphi}{3d_2} \cdot \left( \frac{p_1}{p_2} \right) = 0 \tag{3.2}
\]

where

\[
\hat{d}_1 = d_1(2, y, \varphi^{(1)}, \varphi^{(2)}) \tag{3.3}
\]

\[\text{We note that } \phi \text{ and } \varphi \text{ must be continuously differentiable with respect to } d.\]
is homogeneous of degree zero in \( p \) and \( y \). Then formulate the nonlinear regression model,

\[
d_1^{**} = d_1^* + \varepsilon_1 = d_1^*(p, y, \theta, \varphi) + \varepsilon_1,
\]

(3.4)

where (say)

\[
\varepsilon_1 \sim n(0, \sigma_1^2),
\]

(3.5)

and, since \(-\infty < \varepsilon_1 < \infty\), we also have \(-\infty < d_1^{**} < \infty\). Now define the observed value, \( d_1 \), to "fix-up" the domain, i.e., map all infeasible values into the "closest" feasible boundary value, as in

\[
d_1^{**} = \begin{cases} 
0 & \text{if } d_1^{**} < 0 \\
 d_1 & \text{if } 0 \leq d_1^{**} \leq y/p_1 \\
y/p_1 & \text{if } d_1^{**} > y/p_1.
\end{cases}
\]

(3.6)

This is the standard bilaterally-censored (nonlinear) regression model with p.d.f.,

\[
h(d_1) = \begin{cases} 
 F(0) & \text{if } d_1 = 0, \\
 F(d_1) & \text{if } 0 < d_1 < y/p_1 \\
 1-F(y/p_1) & \text{if } d_1 = y/p_1.
\end{cases}
\]

(3.7)

with \( F \) denoting the p.d.f., \( n(d_1, \sigma_1^2) \). This model can easily be estimated by ML methods (Rosett and Nelson (1975), Hartley and Swanson (1980)), and is portrayed in Figure 1 below:
Finally, we determine \( d_2 \) by the identity,

\[
d_2 = \frac{(y-p_1 \cdot d_1)}{p_2}.
\]

(3.8)

It is worth noting that \( d_1^* \) has an unrestricted domain, and is often specified \textit{a priori} (as a postulated regression function), rather than being derived from an optimization model. In the univariate (2-good) case, this does not necessarily matter, provided the relevant homogeneity restrictions, etc., are imposed. Also, it should be noted that if the inequality constraint,

\[
0 \leq d_1^* \leq y/p_1,
\]
were imposed, as must hold on the observed value, \( d_1 \), then the boundary probabilities, \( \Pr[d_1 = 0] \) or \( \Pr[d_1 = y/p_1] \) cannot exceed 1/2.

2. The Three-Good Case:

We now extend the analysis to the 3-good case, largely as a vehicle to indicate what happens in the D-good situation (as \( D \) increases). Following our paradigm, we use the budget constraint to eliminate (say) the \( D^{th} \) good by substitution. We then solve the first-order (unconstrained) utility maximization problem for \( d_{D-1}^{*}(\varrho, \lambda^{(1)}, \vartheta) \), thus automatically imposing the cross-equation restrictions.  \(^1\) Then, add a \((D-1)\)-dimensional (normal) random vector (say), \( \xi_{D-1} \sim \mathcal{N}(0, \sigma_{D-1}) \), to \( d_{D-1}^{*} \), defining the regression-system,

\[
d_{D-1}^{**} = d_{D-1}^{*} + \xi_{D-1} = d_{D-1}^{*}(\varrho, \lambda^{(1)}, \vartheta) + \xi_{D-1}.
\]

Then, for all infeasible \( d_{D-1}^{**} \) values, "map" them onto the closest boundary point. Then recover \( d_D = (y - \varrho_D d_{D-1}) / \rho_D \).

We exhibit the result of this paradigm for a particular unconstrained solution, \((d_1^{*}, d_2^{*})\), in Figure 2. In this particular case, all possible \((d_1^{*} \xi_1, d_2^{*} \xi_2)\) values on the half-line orthogonal to the line \( d_1 = 0 \) at the observed \( d_2 \) value, with \( 0 < d_2 < y/p_2 \), map into the same observed value, \((0, d_2)\).

\(^1\) This is not done in general, when the system of regression functions, \( d_{D-1}^{*} \) are postulated a priori, as commonly occurs — see, e.g., Madalla (1983).
In Figure 3 we exhibit the regions of integration and line-integrals which result from the "closest-feasible-boundary-value" principle in the three-good case. Note that all vertices, \((0,0)\), \((y/p_1,0)\), and \((0,y/p_2)\), involve mapping an infeasible region in \((d_1,d_2)\)-space onto a point; whereas all other boundary values obtain from line-integrals.

Thus apart from \((d_1,d_2)\) values in the interior, which have the normal p.d.f. (say) \(f(d_1,d_2)\), all boundary points involve numerical integration of \(N(\begin{bmatrix} d_1^* \\ d_2^* \end{bmatrix}, E^*_{2})\) over various regions or half-lines, orthogonal to the relevant boundary line at the observed \((d_1,d_2)\) value, as given below:
Figure 3: The Feasible Set and Boundary Mappings in the 3-good Case

(1) \( d_1 = 0, d_2 = 0 \) with prob. \( \int_{-\infty}^{0} \int_{-\infty}^{0} f(d_1, d_2) \, dd_1 \, dd_2 \)

(2) \( d_1 = y/p_1, d_2 = 0 \) with prob. \( \int_{-\infty}^{y/p_1} \int_{-\infty}^{\infty} f(d_1, d_2) \, dd_1 \, dd_2 \)
   \[ + \int_{0}^{\infty} \int_{0}^{\tan^{-1}(p_2/p_1)} f(r \cdot \sin \omega + y/p_1, r \cdot \cos \omega) \cdot r \, d\omega \, dr \]

(3) \( d_1 = 0, d_2 = y/p_2 \) with prob. \( \int_{y/p_2}^{\infty} \int_{-\infty}^{0} f(d_1, d_2) \, dd_1 \, dd_2 \)
   \[ + \int_{0}^{\pi/2} \int_{0}^{\tan^{-1}(p_2/p_1)} f(r \cdot \sin \omega, r \cdot \cos \omega + y/p_2) \cdot r \, d\omega \, dr \] \( (3.10) \)
(4) \(0 < d_1 < y/p_1, d_2 = 0\) with prob. \(\int_{-\infty}^{\omega} f(d_1, d_2) \, dd_2\)

(5) \(d_1 = 0, 0 < d_2 < y/p_2\) with prob. \(\int_{-\infty}^{0} f(d_1, d_2) \, dd_1\)

(6) \(d_1 > 0, d_2 > 0, y = p_1d_1 + p_2d_2\) with prob.

\[
\int_{0}^{\omega} f(r \sin(\tan^{-1}(p_2/p_1)) + d_1, r \cos(\tan^{-1}(p_2/p_1)) + d_2) \, r \, dr,
\]

where (2b), (3b) and (6) in Figure 3 are obtained by first translating \((d_1, d_2)\) to the origin, and then transforming to polar coordinates.\(^{1/2}\)

\(^{1/}\) It is sometimes argued that, in LDV problems, one way to avoid the "curse of multiple numerical integration" is to assume that the disturbances are independent, so that the joint p.d.f. factors into the product of the marginals. In the present case, this simplifying assumption implies

\[
f(\epsilon_1, \epsilon_2) = f_1(\epsilon_1) \cdot f_2(\epsilon_2) \quad \text{with} \quad \Sigma^* = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.
\]

In inequality-constrained neoclassical models (as in (2.1)-(2.3)), the reader may inspect the expressions for the regions, (2b) and (3b) of (3.10), to verify that this does not avoid the difficulty.\(^{2/}\)

\(^{2/}\) Note that under the model for the traditional approach one may still generate observed values of the decision that are zero, even though this may not be possible under the (deterministic) maintained hypothesis -- as, e.g., with the trans-log direct utility function. Here, "zeroes" are due entirely to random errors.
C. Generalization

Generalization of our simple model of the representative household via the traditional regression-based approach is evidently a formidable task. The number of special cases and the order of numerical integrations proliferate at a rate which quickly exceeds the capabilities of both this investigator and the present state of computer technology. In addition, we regard the underlying paradigm as unsuitable for inequality-constrained, neoclassical economic models. Accordingly, we shall not attempt formulation of the general (D-good) case for the representative household, and leave as an exercise for the reader development of the traditional approach to model the land-use and input-demand system for the profit-maximizing farmer for increasing values of D.

We do, however, note that if, instead, the decision variables were subject to a set of strong inequalities,

\[ d > 0 \]                        \hspace{1cm} (3.11)

i.e., for the representative household, each unit consumes something of each good, then in the 3-good case we would obtain the truncated p.d.f.,

---

1/ See, e.g., Haber (1970), Stroud and Secrest (1972), and Quandt (1983) for discussions of multiple numerical integration.

2/ Such a model may be applicable to aggregated, macro-level time-series data, or to micro-level samples with a very broad classifications of goods and services, etc.
for \(0 < d_1 < y/p_1\) and \(0 < d_2 < (y-p_1d_1)/p_2\). It should be evident that this formulation is \textit{easily} generalized to the D-good case, but does \textbf{not} avoid a \((D-1)^{st}\)-order numerical integration for each \((i,t)\) within each iteration in maximization of the associated log-likelihood function with respect to \(\theta\) and \(\xi\).

IV. A Neoclassical Econometric Approach

As compared with the traditional (regression-based) approach, the neoclassical econometric approach begins with the premise that, in determining their optimal decisions, \(d^*_it\), relative to the maintained hypothesis, (2.1)-(2.3), the inequality restrictions, (2.3), are \textbf{not} ignored.

We shall propose a direct approach to this problem, involving calibration of the parameter vector, \(\theta\), in the context of a \textit{deterministic} model.\(^1\)

\(^1\) Stochastic analogues will be treated in a subsequent paper.
A. A Deterministic Model

The simplest formulation of a neoclassical econometric model is the deterministic case. This model may be stated as

\[ d_{it} = d^*_i + e_{it} = d^*(s_{it}, \theta) + e_{it} \]  

(4.1)

where the decision-vector, \( d^*_i \), is now the solution to the nonlinear/dynamic programming problem, (2.1)-(2.3), with the particular state vector, \( s_i = s_{it} \); and \( e_{it} \) denotes a constant vector of discrepancies between the actual decisions, \( d_{it} \), and the model-solutions, \( d^*(s_{it}, \theta) \), relative to a particular value of \( \theta \). The calibration problem is then to determine a value of \( \theta \) which brings the model-solution values, \( d^*_i \), as "close as possible" to their observed counterparts, \( d_{it} \), associated with the "state of nature," \( s_{it} \), for all \( (i, t) \).

To this end, we adopt (say) the quadratic loss function, \(^1\)

\[ L(\theta) = \frac{1}{2} \sum_{i, t} (d_{it} - d^*_{it})' W_{it} (d_{it} - d^*_{it}) \]  

(4.2)

\[ = \frac{1}{2} \sum_{i, t} e_{it}' W_{it} e_{it} , \]

where, to adjust for different units of measurement in the elements of \( d_{it} \), we employ the weighting matrix,

\[ 1/ \text{ Choice of the form of the loss-function in a deterministic model is at the behest of the investigator.} \]
\[ w_{it} = w(\theta) = \left[ \frac{1}{N \cdot T} \cdot \sum_{i,t} (d_{it} - d^*_{it}) \cdot (d_{it} - d^*_{it})' \right]^{-1}. \] (4.3)

The problem is to determine a value of \( \theta \), which minimizes \( L(\theta) \) of (4.2).

B. A Simple Algorithm

Let \( \theta^{(0)} \) denote a feasible initial value for the parameter vector, \( \theta \), i.e., a value such that, for every \( (i, t) \) within the data set, a solution, \( d^*_{it} \), may be computed as a solution to the nonlinear/dynamic programming problem, (2.1)-(2.3), for \( \theta = \theta_{it} \). Let \( n = 1, 2, \ldots \) denote an iteration index; and for any function, \( u(\theta) \), let \( u^{(n)}(\theta) \) denote the arbitrary n\(^{th}\) iteration:

**Iteration n:**

The "Inner (Optimization) Algorithms":

**Basic Solution:** Given \( \theta^{(n)} \), for every \( (i, t) \) within the data set, determine the solution-vector, \( d^*_{it}^{(n)} \), to the programming problem:

"Maximize

\[ g(d, z_{it}, \theta^{(n)}) \]

subject to \( g(d, z_{it}, \theta^{(n)}) \)

and

\[ d \geq 0. \]

This permits evaluation of the loss-function, \( L^{(n)}(\theta) = L(\theta^{(n)}) \).
Perturbed Solutions: We must now determine a direction and step-size, relative to the feasible parameter vector, $\hat{g}^{(n)}$, which (locally) reduces the loss-function, $L(\hat{g})$, relative to the prevailing value, $L^{(n)}$. Here, we have an inequality-constrained set for the domain of $d$, defined on $S^0 \times O^0$ under the maintained hypothesis. Also, $d_{it}^*(n) \equiv d_{it}^*(d_{it}, \hat{g}^{(n)})$ is feasible for every $d_{it} \in S^0$, relative to $\hat{g}^{(n)} \in O^0$. The problem, therefore, is to determine an update of the parameter vector, say $\hat{g}^{(n+1)}$, such that:

1. $d_{it}^*(n+1)$ is feasible for every $(i,t)$, and
2. $L^{(n+1)} \leq L^{(n)}$.

Consider the first-order conditions,

$$\frac{\partial L(\hat{g})}{\partial \hat{g}} = \sum_{i,t} J_{it} \cdot \hat{W}_{it} \cdot (d_{it} - d_{it}^*)$$

$$= J \cdot \text{Diag}(W) \cdot (\hat{g} - d_{it}^*) = 0,$$  \hspace{1cm} (4.4)

where $J = (\partial d/\partial \hat{g})$ is a $D \times N \times T \times P$ Jacobian matrix; $\text{Diag}(W) = \text{Diag}(\hat{W}_{it})$; $d = \text{vec}(d_{it})$ and $d^* = \text{vec}(d_{it}^*)$. We employ quasi-linearization methods (Bellman and Roth (1983)), to iteratively replace $d^*$ in (4.4) by a linear approximation around $d_{it}^{(n)}$, i.e.,

$$d^* \approx d_{it}^{(n)} + J_{it}^{(n)} \cdot (\hat{g} - \hat{g}^{(n)}).$$  \hspace{1cm} (4.5)

while holding fixed $J$ and $W$ at $\hat{g}^{(n)}$. Solution for $\hat{g}$, using the iterative approximation, (4.5), leads to the parameter updating condition,
\[ \phi(n+1) = \phi(n) + \lambda(n) \cdot [j(n)' \cdot \text{Diag}(W(n)) \cdot j(n)]^{-1} \cdot [j(n)' \cdot \text{Diag}(W(n)) \cdot (d-d^*(n))], \quad (4.6) \]

where \( \lambda(n) \) is a scalar step-size parameter.\footnote{See Dennis and Schnable (1983) for a comprehensive discussion of line searches, scaling and stopping rules for such "quasi-Newton" methods.}

In the present case, the problem is how to define the elements of 
\[ j(n) = (3d^*(n)/3\theta_p) \text{ in the parameter updating condition, (4.6).} \]

Since all elements of \( d^*(n) \) are, by construction, non-negative, i.e.,
\[ d^*_i \geq 0 \text{ for all } (i,t); \text{ and some elemen } i, \ d^*_i = 0, \text{ may be at binding values, relative to } \phi^{(n)}, \text{ evidently the finite-difference approximations for} \]
\[ \frac{3d^*_i}{3\theta_p}, \ p = 1, \ldots, P, \text{ vary according to the case at hand: Hence,} \]

\[ \frac{3d^*_i}{3\theta_p} = (d^*_i + d^*_i(p,n)) \cdot \delta, \quad (4.7) \]

where the '-' sign and a positive perturbation, \( d^*_i = d^*_i(\phi^{(n)} + \delta \cdot i_p) \) is employed, if \( d^*_i \) is outside a \( \delta \)-neighborhood of the boundary, \( d = 0 \); and where the '+' sign and a negative/positive perturbation, 
\[ d^*_i = d^*_i(\phi^{(n)} + \delta \cdot i_p) \text{ is employed, if } d^*_i \text{ is } \text{within a } \delta \text{-neighborhood of the boundary of the feasible decision-space. Here, } \delta > 0 \text{ is a suitably-small perturbation in the } p \text{th element, } \phi_p^{(n)}, \text{ of } \phi^{(n)}; \text{ and } i_p \text{ is the zero vector, except for a unit in the } p \text{th position. In short, the proximity to the boundary and the (local) positivity/negativity of the first derivative at issue determines the nature of the (one-sided/two-sided) parameter perturbation, } p = 1, \ldots, P, \text{ in } \phi^{(n)} \text{ to be employed; where each case can be separately treated and verified numerically. We thus obtain:} \]
The "Outer (Parameter Updating) Algorithm": Calculate \( \dot{f}^{(n)} = \text{Vec}(\partial d^{(n)}/\partial \vartheta^{(n)}) \) via the \( p^{(n)} \), \( 2P \geq p^{(n)} \geq P \), perturbations in the elements of the \( P \times 1 \) parameter vector, \( \dot{\vartheta}^{(n)} \), and determine \( \vartheta^{(n+1)} \) via (4.6) for suitable direction, scaling and step-size procedures in our "quasi-Newton" algorithm.\(^1\)

C. Discussion

Our simple algorithm thus involves embedding a set of "inner" nonlinear/dynamic programming problems -- one for every \((i,t)\) within the data set -- within an "outer" parameter updating algorithm which searches for a new value of \( \vartheta \) which reduces the loss-function value, while retaining feasibility of all decisions. The overall algorithm involves iterating between the set of "inner" and "outer" algorithms until convergence to a limit point,

\[ \hat{\vartheta} = \lim_{n \to \infty} \{ \vartheta^{(n)} \}, \]

is achieved.

Insofar as our model is deterministic, we must therefore forgo the possibility of performing customary "statistical" tests of hypotheses and other inferences, i.e., by treating \( \hat{\vartheta} \) (and \( \hat{\xi} \)) as estimates of \( \vartheta \) (and \( \xi \)) and regarding \( \vartheta \) as the maintained hypothesis, (2.1)-(2.3), as the "true" parameter value. In return, we believe our approach to calibration rests upon a more plausible paradigm, and thus should determine values of \( \vartheta \) more consistent with the maintained hypothesis.

The algorithm proposed is by no means the only feasible numerical approach. In many applications, it will be possible to utilize penalty functions to account for the inequality constraints. One ingenious suggestion...

\(^1\) See Dennis and Schnabel (1983) for a comprehensive discussion.
(due to George Box -- see Dixon (1972), p.100) involves replacing the maximand, \( g(d, s_{it}, g(n)) \), in the "inner algorithms" by the functions,

\[
g^*(d, s_{it}, g(n)) = g(d, s_{it}, g(n)) - D \sum_{j=1}^{d_j} d_j^{2}H(d_j),
\]

where \( H \) is the Heaviside step-function,

\[
H(d_j) = \begin{cases} 
0 & \text{if } d_j \geq 0 \\
1 & \text{if } d_j < 0
\end{cases}
\]

and \( \rho > 0 \) is a very small number.\(^1\) Alternative forms of penalty functions may also be employed to treat the case of strong inequality constraints, as in (3.11), within the present framework.\(^2\)

Finally, the case of arbitrary configurations of incomplete data in the \( \{d_{it}, s_{it}\} \) may be treated by the methods of Hartley (1984c, 1985a, 1985b).

\(^1\) Choice of \( \rho \) is critical, as the problem becomes "ill-conditioned" when \( \rho \) is "too small".

\(^2\) We look forward to accumulating some computational experience on the various algorithms which may be employed within our overall neoclassical econometric approach.
V. Conclusions, Extensions and Future Research

A. Conclusions

This paper has considered a class of neoclassical economic models in which the representative decision-making unit of a given type is presumed to solve a constrained optimization problem, involving both equality and non-negative (weak/strong) inequality restrictions. We have provided two simple illustrations of such models and discussed how the canonical form may be solved via nonlinear and/or dynamic programming algorithms provided the parameters in the primal behavioral/technological (utility/production) functions are known. We then turn to the inverse problem -- i.e., how to determine the parameter values from a sample/set of data on the decision and state variables.

The traditional (regression-based) approach is first considered. By analogy with the statistical foundations of the single-equation, censored-regression model, in which all inadmissible decision values are mapped onto the closest boundary value (e.g., Tobin (1958), Amemiya (1973, 1984), Rosett and Nelson (1975), Hartley (1976), Hartley and Swanson (1980), etc.), we formulate a general paradigm for the multiple-decision case -- arising from a neoclassical maintained hypothesis. Here, we find that, in order to avoid restricting the probability with which boundary solutions will obtain, while at the same time imposing cross-equation restrictions on the resulting set of decision functions, the traditional approach is forced to adopt an apparently implausible paradigm. Under our characterization, decision-makers are presumed to: (1) solve an equality-constrained optimization problem in which all inequality domain restrictions are ignored; (2) add a multivariate (normal) disturbance to the functionally-independent decision sub-vector (as
determined in (1)); and (3) then impose the inequality restrictions by
mapping, *ex post*, all infeasible values onto the closest possible feasible
point on the boundary.

Step (1) imposes cross-equation restrictions on the parameters of the
decision-functions -- but by ignoring the inequality constraints, and hence
adopts an implausible definition of the "location" of the regression-system
mean. In addition, the generalization to multiple-decision problems requires
that the investigator derive the form of a multitude of line and arc
integrals, the evaluation of which are, at present, beyond computer
capabilities. This suggests we look for a simpler, more plausible
formulation.

Our deterministic neoclassical econometric approach, though highly
computer-intensive, appears more directly related to the maintained
hypothesis, and avoids the investigator having to derive the form of and
compute various multiple integrals inherent in the traditional methodology.

Given an initial (feasible) parameter vector, our method involves
iterating between a set of "inner" nonlinear and/or dynamic programming
problems -- one for each sample member, relative to the prevailing model
calibration -- and an "outer", parameter-updating algorithm, which searches
for an improved calibration of the model, such that the solution values are
brought "closer" to the observed decisions. Iteration between the "inner" and
"outer" algorithms proceeds until convergence obtains.

As the present neoclassical econometric model is deterministic, we
must sacrifice the various aspects of "statistical" inference -- hypothesis
testing, confidence limits, etc. -- and rely, instead, upon "sensitivity
analysis" of the assumptions under the maintained hypothesis (see, e.g.,
Bellman and Dreyfus (1962)). In return, we believe the investigator benefits
both in being able to calibrate more realistic models, and by letting the computer do all of the analysis.

In this context, it is appropriate to enquire whether, under the present state of computer technology, our methods are both feasible and cost-effective, when applied to problems involving many decisions, many state variables and many sample members. We look forward to obtaining access to a supercomputer and reporting on such attempts in a subsequent paper.

B. Extensions and Future Research

We have argued that the stochastic formulation of the model, (2.1)-(2.3), as given by the traditional (regression-based) approach is implausible — particularly in a multiple-decision context. The first extension, therefore, is to formulate plausible stochastic analogues to our deterministic maintained hypothesis. We shall argue (in a subsequent paper) that in order to preserve the essential nonlinear/dynamic programming structure of the maintained hypothesis, which permits calculation of a numerical solution, \( \hat{d}^* \), to a deterministic problem, the stochastic features should be introduced directly into the objective function and/or equality constraints. This involves treating certain of the state variables, \( \mathbf{s} \), or parameters, \( \mathbf{\theta} \), as random variables (e.g., Hartley (1983a, 1983b)), or, otherwise, introducing unobserved random errors directly into the objective function (e.g., Bellman and Dreyfus (1962), McFadden (1974), Wolpin (1982)).

The difficulty, of course, is that since an analytic, closed-form solution for \( \hat{d}^* \) is impossible, except in special cases such as that of McFadden (1974) — even in the deterministic case, it will not be possible to derive an expression for the p.d.f., \( p(\hat{d}^*(\mathbf{s}(1),\mathbf{\theta}(1),\mathbf{\theta})) \) regardless of which
scenario is adopted for a stochastic analogue; where \( \mathbf{s}^{(1)} \) denotes the observed sub-vector of state-variables, \( \mathbf{a}^{(1)} \) denotes the constant sub-vector of parameters and \( \mathbf{a} \) denotes the parameters in the p.d.f. of \( \mathbf{s}^{(2)}, \mathbf{a}^{(2)} \) and \( \xi \). This suggests we consider numerical methods.

Our proposal is to employ "Iterated Monte Carlo" procedures, \(^1\) involving repeated sampling from a postulated p.d.f. (say) \( h(s^{(2)}, a^{(2)}, \xi; a) \) with "known" parameter vector, \( a^{(n)} \), to numerically construct the p.d.f. of \( a^{(n)} \) above. Then by repeated perturbations in the elements of \( a^{(n)} \), we can approximate the derivative of the likelihood function, using methods analogous to Hartley (1981a, 1983b, 1984c). Needless to say, the computational cost of such methods is enormous, at present. Accordingly, to preserve the realism of the model structure, our preference is to begin with deterministic models; and to introduce stochastic features parsimoniously, and only where important. Details will be left to a subsequent paper -- see, however, Hartley (1983b).

The next extension is to consider more general types of domain restrictions -- e.g., integer-valued decision variables, general convex sets, etc. In these situations, the "inner algorithms" becomes integer-programming (e.g., Gomory (1963)), mixed-continuous-integer programming (e.g., Beale (1958)) or, in general, dynamic programming (Bellman (1957)) problems. The problem is how to formulate an appropriate "outer" parameter-updating algorithm, whether in the deterministic or stochastic case, i.e., how to select the form of the appropriate parameter perturbations. Here, similar methods should be applicable.

---

\(^1\) This approach was sketched in Hartley (1984a), and mis-labeled a "Bootstrap Monte Carlo" method -- a term already used by Efron.
Our analysis, at this juncture, has implicitly assumed a partial equilibrium model. In a competitive general equilibrium model, the prices (which have been treated as exogenous state variables in a partial equilibrium setting) now become non-decision endogenous state variables from the perspective of the individual economic unit. Thus, the decisions of each unit are no longer modeled as independent of those of its neighbors. Aggregation of the demand/supply decisions of individual households/firms/etc., in a market system, determine all (relative) prices, which, in turn, feed back (in endogenous fashion) to determine such decisions. Calibration methods, employing the "systems approach" of Hartley (1984, 1985a) have been proposed to treat this problem, in the context of arbitrary configurations of available data.

Finally, the maintained hypothesis must be generalized to include dynamic, multi-stage optimization problems. For examples, see Wolpin (1982), Rust (1984), and Bellman and Hartley (1985).


Hald, A. (1949), "Maximum Likelihood Estimation of the Parameters of a Normal Distribution which is Truncated at a Known Point," Skandinavisk Aktuarietidskrift, 32, 119-134.


