

Can Subjective Questions on Economic Welfare Be Trusted?

Evidence for Three Developing Countries

Martin Ravallion

Kristen Himelein

Kathleen Beegle

The World Bank
Development Research Group
Poverty and Inequality Team
December 2013



Abstract

While self-assessments of welfare have become popular for measuring poverty and estimating welfare effects, the methods can be deceptive given systematic heterogeneity in respondents' scales. Little is known about this problem. This study uses specially-designed surveys in three countries, Tajikistan, Guatemala, and Tanzania, to study scale heterogeneity. Respondents were asked to

score stylized vignettes, as well as their own household. Diverse scales are in evidence, casting considerable doubt on the meaning of widely-used summary measures such as subjective poverty rates. Nonetheless, under the identifying assumptions of the study, only small biases are induced in the coefficients on widely-used regressors for subjective poverty and welfare.

This paper is a product of the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at kbeegle@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Can Subjective Questions on Economic Welfare Be Trusted?

Evidence for Three Developing Countries

Martin Ravallion, Kristen Himelein and Kathleen Beegle¹

Key words: Subjective welfare, subjective poverty, scales, heterogeneity, vignettes

JEL classifications: C81, D60, I32

Sector board: Poverty Reduction (POV)

¹ Ravallion is with Georgetown University, Beegle and Himelein are with the World Bank. This paper would not have been possible without the work of the *Tajikistan* Living Standards Survey Team, including Diane Steele, Sasun Tsirunyan, Vladimir Kolchin, Farhod Khamidov, Oleksiy Ivaschenko, and staff of the National Statistics Committee of Tajikistan (*Goskomstat*), Economic Development Initiatives in Bukoba, Tanzania, including Joachim De Weerd and Albertus Kamanzi, Centro de Estudios y Documentación de la Frontera Occidental in Huehuetenango, Guatemala, and the Programa Regional de Seguridad Alimentaria y Nutricional para Centroamérica for funding the Guatemala portion of fieldwork. Helpful comments on this paper were received from Denis Cogneau, Sylvie Lambert and Claudia Senik. All views are those of the authors and do not reflect the views of their employers including the World Bank or its member countries.

1. Introduction

Widely used measures of subjective welfare ask survey respondents to rate their “economic welfare,” “satisfaction with life” or “happiness” on an ordinal scale. These measures have found innumerable applications in the psychological and social sciences and have recently become popular in economics.² However, different people may well have different ideas about what it means to be “rich” or “poor,” or “satisfied” or not with one’s life, leading them to interpret survey questions on subjective welfare differently.³ For example, the Young Lives Project (2009) reports the comment of a six-year old in rural Vietnam, named Duy, as saying that “We are nearly rich as we have a new cupboard, but we haven’t got a washing machine.” Duy clearly has a different idea of what it means to be “rich” than those in Vietnam more familiar with the living conditions of the truly rich. Survey respondents can be expected to interpret subjective questions relative to their personal frame-of-reference, which will depend on latent aspects of their own knowledge and experience.

Two important applications of subjective welfare data illustrate why this matters. The first is their application in the interpersonal comparisons of welfare required for poverty measurement. Measures of “subjective poverty” are becoming common.⁴ These measures tell us what proportion of survey respondents place themselves on the bottom rung (or possibly second lowest rung) of a welfare ladder from “poor” to “rich.” But if the rungs of the welfare ladder are not understood the same way by different respondents it is unclear what meaning can be attached to such measures.

The second application relates to the many studies of the covariates of subjective welfare.⁵ In now standard practice, a linear or ordered probit (OP) regression is run of the survey

² The relevant economics literature is reviewed by Frey and Stutzer (2002), Di Tella and MacCulloch (2006) and Dolan et al. (2008). The psychological literature on subjective welfare is reviewed in Diener et al. (1999) and Furnham and Argyle (1998). An alternative approach is to ask what level of income is needed to attain a given position on a ladder, such as not being “poor.” This is the “Leyden method” devised by van Praag (1968). While we do not use this type of data here, the same concerns about bias arise in the Leyden method.

³ While this paper focuses on heterogeneity in scales, there are other concerns with survey design. For example, Conti and Pudney (2011) find that minor re-designs in questions on satisfaction of life/work led to large changes in answers, particularly for women, finding that distortions in responses influence findings with respect to correlates of women’s job satisfaction. For an overview of the concerns about inferring welfare effects from subjective data see Ravallion (2012).

⁴ Examples include Mangahas (1995), Ravallion and Lokshin (2002), Carletto and Zezza (2006), and Posel and Rogan (2013).

⁵ Examples include van de Stadt et al. (1985), Clark and Oswald (1994, 1996), Kapteyn et al. (1998), Oswald (1997), Winkelmann and Winkelmann (1998), Pradhan and Ravallion (2000), McBride (2001), Ravallion and

responses against individual and household characteristics, such as age, gender, marital status, income, education, employment status, and household demographics. Such regressions offer the prospect of identifying various welfare effects and trade-offs of interest (including to policy makers) under seemingly weaker identifying assumptions than required by widely-used methods that rely solely on objective circumstances, such as income or consumption. We can agree in principle that a person's economic welfare does not only depend on the household's current consumption or income, but is also influenced by the size and demographic composition of the family and characteristics such as education and employment. "Prices" are missing for these other attributes. Subjective data offer a solution for identifying the trade-offs and constructing a composite index of welfare based on the regression's predicted values.

But can that solution be trusted? The OP estimator assumes that the thresholds—the values of the underlying welfare metric at which ordinal responses on the stipulated scales change—are constant parameters, the same for all respondents. We define "scale heterogeneity" as any situation in which this assumption does not hold, i.e., that the thresholds are idiosyncratic. If there is such heterogeneity and it is correlated with the covariates in subjective welfare regressions then biased inferences about the underlying welfare function will be drawn from the regressions found in the literature. (This concern arises in addition to more familiar concerns about the possible endogeneity of regressors, which create confounding correlations with the error term in the underlying continuous variable for subjective welfare.)

Concerns about such systematic measurement errors in subjective questions have prompted some observers to warn against their use as dependent variables. Bertrand and Mullainathan (2001, p.70) conclude that:

"...subjective variables cannot reasonably be used as dependent variables, given that the measurement error likely correlates in a very causal way with the explanatory variables."

This dismisses a great many past and potential applications using subjective welfare questions. But is such a negative assessment really warranted?

It would be fair to say that the potential problem of systematic scale heterogeneity has received little more than passing attention in the extensive empirical literature making subjective welfare comparisons. Dolan et al. (2008) survey the findings of a large number of papers running

Lokshin (2001, 2002, 2010), Graham and Pettinato (2002), Senik (2004), Luttmer (2005), Ferrer-i-Carbonell (2005), Graham and Felton (2006), Herrera et al. (2006), Bishop and Luo (2006), Kingdon and Knight (2006, 2007), Fafchamps and Shilpi (2009), Knight and Gunatilaka (2010, 2012), Castilla (2010) and Posel and Rogan (2013).

regressions for subjective welfare but do not explicitly discuss the potential for bias due to systematic differences in scales (though they do note concerns about the possible endogeneity of some regressors).⁶ A seemingly widely-held view is reflected in the authoritative survey paper by Frey and Stutzer (2002), which notes the scope for scale heterogeneity in self-reported welfare responses but claims that this does not invalidate regression models for such data. That claim is hard to defend on *a priori* grounds given the aforementioned concerns about bias.

It would seem premature either to ignore the problem (following the advice of Frey and Stutzer) or to abandon subjective poverty/welfare regressions knowing only that there is a potential for bias (following Bertrand and Mullainathan). More evidence is needed.

Asking survey respondents to place vignettes describing hypothetical situations on the same scale has been used to address scale heterogeneity in a few studies of subjective data on health status, political efficacy, and job satisfaction.⁷ Following this approach, Beegle, Himelein and Ravallion (BHR) (2012) used vignettes to study frame-of-reference effects on subjective welfare and offered various tests for confounding effects of scale heterogeneity using data for Tajikistan.⁸ BHR found that, despite the existence of scale heterogeneity, subjective welfare regressions that ignored the problem still gave quite similar results to those that address it.

The present paper makes two contributions, corresponding to the two applications described above. First, we propose a new measure of subjective poverty, anchored to a household vignette designed for describing a family that most would consider poor in the specific setting. Instead of counting as poor those who put themselves on (say) the lowest rung of a welfare ladder, with uncertain meaning and comparability, subjective poverty is measured by comparing the household's self-assessed welfare to its assessment of the welfare of this specific poor household described in the vignette. In other words, we ask for explicit inter-personal comparisons of welfare against a common reference. The paper discusses this approach in theoretical terms and provides applications to three developing countries, Tajikistan, Guatemala

⁶ Some papers run linear regressions for the ordinal responses on subjective welfare rather than an OP. The assumption of constant scales is explicit in the OP but the problem is clear still present in the linear models.

⁷ King et al. (2004) and King and Wand (2007) designed vignettes to establish common points on the heterogeneous reference scales regarding political efficacy in China and Mexico. Kristensen and Johansson (2008) used vignettes in anchoring subjective scales for job satisfaction. Kapteyn et al. (2008) use vignettes to compare life satisfaction between respondents in the U.S. and the Netherlands. Bago d'Uva et al. (2008) used them for correcting self-assessed health data for reporting bias.

⁸ This can be thought of as a contribution to ongoing efforts to employ qualitative data to help validate standard "objective" metrics of welfare and poverty. For an overview of various approaches see Shaffer (2013).

and Tanzania. In studying the scale-heterogeneity problem in relatively poor settings we can reasonably expect to obtain a more complete characterization of living standards than would be possible with a short vignette in a rich country. Related to our new measure of subjective poverty, we propose a measure of scale heterogeneity among the poor. We find considerable scale heterogeneity and substantially higher poverty rates with our new measure, but the empirical determinants of subjective poverty turn out to be very similar to past methods ignoring scale heterogeneity.

Second, the paper tests the robustness of the conclusions of BHR with regard to the extent of biases in standard regression models for subjective welfare. Here we follow what appears to be the most common approach in the literature, whereby the survey responses are interpreted as ordinal indicators of a latent continuous welfare metric. With the additional assumptions of constant thresholds (the levels of welfare at which ordinal responses switch along the scale) and a normally distributed error term in the latent welfare variable, an OP is then widely used to model the data to retrieve the parameters of the underlying welfare function. In common with past work, the scales identified in the survey question are not treated as having any welfare significance—they are merely the survey instrument used to help identify the underlying welfare metric. The difference with past work is that we use the vignette scores to relax the assumption that the thresholds in the welfare space are constant across respondents. BHR also addressed this issue but only had the vignette data for Tajikistan, for which they found that respondents with different socioeconomic backgrounds tended to use systematically different scales in responding to subjective welfare questions. This prompted us to investigate the issue elsewhere. We decided to pick two very different poor areas of the world, in Guatemala and Tanzania. In both cases the study areas are clearly poor, but they are not unusually isolated or equal, so it can be expected that people will have some knowledge of the range of living standards in their societies. Using the vignettes developed for this study in these quite different settings, we confirm the finding of BHR that subjective welfare regressions are reasonably robust to scale heterogeneity. These findings put applications using subjective poverty and economic welfare data as dependent variables on a firmer foundation.

We begin with a description of our data. Section 3 presents our approach to measuring subjective poverty using vignettes. That section also tests whether different covariates emerge,

compared to past measures of subjective poverty. Section 4 turns to our tests of the robustness of regressions for subjective economic welfare. Section 5 concludes, also noting some caveats.

2. Survey data on subjective economic welfare

We study subjective economic welfare, as measured by survey responses to the following question: *“Imagine a 6-step ladder where on the bottom, the first step, stand the poorest people, and the highest step, the sixth, stand the rich. On which step are you today?”* Concerns about scale heterogeneity also arise with questions related to broader welfare concepts such as “happiness” or “satisfaction with life.” Our methods may well be adapted to these broader concepts, although it is likely to be harder to devise credible and practical vignettes for such questions, given that so many variables could be deemed relevant.

Respondents were asked to place themselves on the subjective welfare ladder described above. Later in the questionnaire they were asked to place four vignettes, each describing a hypothetical household, on the same ladder, and finally to (again) place their own household on the ladder, after scoring the vignettes. In asking the own-welfare questions both before and after the questions about the vignettes, we are able to test whether the vignettes alter the respondents’ perceptions of their welfare. The vignette questions may focus the respondent to think about, and possibly revise, the scale they have in mind in reporting their subjective welfare (similarly, see Hopkins and King, 2010).

For all three countries, we developed the vignettes in consultation with local counterparts. The actual vignettes from the questionnaires (translated into English) are given in the Appendix. The vignettes were designed to capture representative snapshots of various levels of welfare in each country. The first vignette was designed to present a scene that almost anyone in the country concerned would deem to be one of poverty. This was not intended to be the poorest imaginable destitute household, but rather an undeniably poor family with a sustainable livelihood. The second vignette indicated a family with improved conditions, though a family that some would still consider poor in that setting. The third was intended to represent a family from the middle class and the fourth an affluent family. The characteristics incorporated in the vignettes varied across countries and included land holdings, education, diet, clothing, and the ability to heat the home during the winter. The vignettes were developed in a clear expected hierarchy of dominance with respect to economic welfare, with all aspects of socio-economic

status increasing monotonically. This structure was used to minimize the effects of multi-dimensionality, which can lead to the perverse sequencing of the vignettes with respect to the intended ordering if respondents place different values on various characteristics contained in the vignettes. Such perverse rankings are inconsistent with our theoretical model (outlined below), which implies that respondents will agree on the ordering of the vignettes. This does not seem to be a concern for our vignettes since there were very few instances of “incorrect” ordering.⁹ The most common characteristic of respondents who perversely order the vignettes is a low level of education of the household head. These cases of incorrect ordering were excluded from the analysis.

In Tajikistan, the subjective welfare experiment was embedded in the 2007 *Tajikistan Living Standards Measurement Survey* (TLSMS). The sample consists of 4,860 households interviewed in September-November 2007. The sample is designed to be representative at the national, urban and rural levels, and at the *oblast* (administrative region) level. Data were collected in two visits. In addition to the standard questions common to multi-topic household questionnaires, subjective welfare modules were developed and asked in the first visit.

In Guatemala, the *Impacto de las Remesas y la Migracion sobre la Seguridad Alimentaria y Nutricional* (IRMSAN) survey was done in the department of Huehuetenango, Guatemala, located in the Western Highlands and bordering the Mexican state of Chiapas. This area is entirely rural and is characterized by high levels of both poverty and temporary outmigration. The survey has a sample size of 1,222 households interviewed from April-August 2008. The survey covered six micro-regions within the municipalities of Cuilco, San Gaspar Ixchil, Santa Ana Huista, and Jacaltenango—selected purposively to capture geographical heterogeneity. The survey was stratified to over-sample migrant households. Four months prior to fieldwork, a household census was conducted in the region. Probability weights are used in the analysis to adjust for the differing probabilities of selection.

In Tanzania, the *Kagera Subjective Welfare Survey* was purposively designed for evaluating the potential use of anchoring vignettes in subjective well-being measurements. Fielded in November-December 2007, the survey consists of 450 households randomly selected from Ngara district of Kagera region on the border with Rwanda. This is a poor area in Tanzania

⁹ In Tajikistan, 89 of the 4,860 households in the sample had incorrect coding. In Guatemala the corresponding number was 28 out of 1,222 and in Tanzania it was 3 out of 450.

but also an area where infrastructure (especially roads) has improved in recent times (after the genocide in Rwanda). This has opened up the Kagera area, including the introduction of information about living standards elsewhere.

While the Tajikistan and Guatemala surveys were entirely verbal, in the case of Tanzania, the survey team used a paper diagram of the “ladder of life” onto which respondents placed blocks for each of the household vignettes and their own household. This was developed after piloting showed that verbal-only implementation did not work well. Half-steps were also allowed (scoring at, for example, 2.5).

In Table 1 we compare the pre- and post-vignette responses of the household’s subjective welfare. While there are some off-diagonal elements, the correspondence is quite strong. Most respondents place themselves in the same position in the pre- and post-vignette. (The footnote to each panel of the table gives summary statistics on the strong correlation in the cross-tab.) In all three cases, the modal responses are around the second or third rung of the ladder. Notice that very few respondents put themselves on the top rung of the ladder; pre-vignette the proportion is only 0.2%, 0.1% and 0.4% for Tajikistan, Guatemala and Tanzania respectively. This could well reflect an unwillingness of the rich to reveal their true economic position.

3. Subjective poverty measures

The vignettes allow us to make explicit at least a sub-set of the characteristics that define “poverty.” This is in contrast to the standard approach to measuring subjective poverty in which no characteristics are explicit.

3.1 Theoretical representation

Let SW_i denote the subjective welfare of respondent i in a sample of N respondents. This is a continuous but latent variable. The survey provides an ordinal response R_i on subjective welfare, which is assumed to be increasing in SW_i as follows:

$$R_i = 1 \text{ if } SW_i < \tau_i^1 \text{ for } i=1,\dots,N \quad (1.1)$$

$$R_i = k \text{ if } \tau_i^{k-1} \leq SW_i < \tau_i^k \text{ for } i=1,\dots,N \text{ and } k=2,\dots,K \quad (1.2)$$

where the thresholds are $\tau_i^1 < \tau_i^2 < \dots < \tau_i^K$ for the K possible ordinal responses (with $K=6$ in our case). In standard applications of subjective welfare data the thresholds are taken to be fixed across all i while here we take them to be heterogeneous.

The standard subjective poverty measure designates a respondent to be “poor” if $SW_i < \tau_i^1$ ($R_i = 1$). Using this approach we see in Table 1 that 7.5%, 25% and 17% of the samples in Tajikistan, Guatemala, and Tanzania respectively are deemed to be poor (post-vignettes).

However, such measures lack concrete meaning in terms of living standards when the scales vary across respondents. With scale heterogeneity, the resulting subjective poverty measures need not be welfare consistent in that rankings across different respondents in terms of their R_i 's need not accord with their rankings in terms of SW_i . Someone responding that $R_i = 2$ (say) could well have a lower value of SW_i than someone responding that $R_i = 1$. Plainly this could only be ruled out on *a priori* grounds if the scales are constant.

The vignettes help address this problem by fixing a set of welfare-relevant characteristics so as to limit the variance in the underlying idiosyncratic scales. We assume that SW_i is a stable function of a vector of variables, Z , so we can write $SW_i = SW(Z_i)$. However, Z need not be fully observable. The vignette characteristics are interpreted as a subset of the characteristics in the vector Z . We can partition that vector as $Z = (Z^v, Z^o)$ where Z^v is the vector of characteristics identified in the vignettes and Z^o are other characteristics. Z^v can take various values, denoted Z_j^v where $j=1, \dots, J$, which define each of the J vignettes, where in our case $J=4$. The values taken by $SW(Z_j^v, Z^o)$ for vignette j at any given Z_j^v are treated as a random variable, the distribution of which reflects the respondent's uncertainty about the omitted characteristics, Z^o .

We make two key identifying assumptions. First we assume internal consistency between vignette and own-welfare assessments. Specifically, it is assumed that each respondent uses the same subjective welfare function in assessing the welfare of the vignettes as for the respondent's

own welfare and that each respondent uses the same scales (albeit personal scales) when assessing own-welfare and the vignette welfare.¹⁰

Secondly, we assume that each respondent's ordinal responses on subjective welfare of each vignette, denoted R_i^j , are generated by an underlying continuous variable given by their expected values of the welfare of that vignette and that respondents share a common distribution of the unobserved vignette characteristics, Z^o i.e.,¹¹

$$SW^j = E[SW(Z^v, Z^o) | Z^v = Z_j^v] \text{ for } j=1, \dots, J \quad (2)$$

Under these assumptions, SW^j is a stable function of the vignette characteristics and responses in scoring vignette welfare are given by:

$$R_i^j = 1 \text{ if } SW^j < \tau_i^1 \quad (3.1)$$

$$R_i^j = k \text{ if } \tau_i^{k-1} \leq SW^j < \tau_i^k \text{ for } k=2, \dots, K \quad (3.2)$$

Thus, even though the thresholds vary across respondents, all will agree on the ordering of the vignettes.

Probably the greatest concern about these identifying assumptions is the possibility of heterogeneity in the distributions of the omitted characteristics in each vignette. This could yield disagreements among participants in the ordering of the vignettes. However, as noted in the previous section, this was rare empirically; 98% in Tajikistan and Guatemala and over 99% in Tanzania gave the same, expected, ordering of the vignettes.

With this set-up, our proposed vignette-consistent measure of subjective poverty says that respondent i is “poor” if (and only if) $R_i \leq R_i^1$ i.e., the respondent is no better off than the poorest vignette. (More generally, one can choose any reference vignette $r \geq 1$ and say that a person is poor if $R_i \leq R_i^r$.) This still does not guarantee that anyone deemed to be poor by this criterion has a lower $SW(Z_i)$ than those not deemed to be poor. That would require that the vignettes provide a complete description of the welfare-relevant characteristics. The uncertainties about those characteristics and the practicalities of surveying (notably the length of the questionnaire) will

¹⁰ The latter part of our assumption is identical to the assumption of “response consistency” in King et al. (2004).

¹¹ One can relax the latter assumption to allow for additive idiosyncratic differences in the level of subjective welfare assigned to a given vignette. This error term can then be subsumed in the τ_i^k 's. Consistency requires that the same additive error term appears in own-welfare.

undoubtedly preclude such completeness. However, our model does imply a form of consistency in terms of expected welfare, in the specific sense that $R_i < R_i^1$ if (and only if) $SW_i < SW^j$.

Equality of the ladder rungs ($R_i = R_i^1$) does not, of course, assure that $SW_i = SW^j$. In that respect, our measure is no different to standard measures of subjective welfare in that it is purely ordinal, while the underlying levels of subjective welfare are taken to be continuous.

The vignette-consistent subjective poverty rate can be no lower than the ordinary raw poverty rate (the proportion with $R_i = 1$). The vignette-consistent rate will pick up all those who put themselves on the lowest rung. But it will also include those who think vignette 1 is not in fact on the lowest rung and yet still rate their own household at or below that vignette. Clearly such people have a very different frame-of-reference and perceive the existence of even greater deprivations than described by vignette 1. The difference between the poverty count based on $R_i = 1$ and that based on $R_i \leq R_i^1$ reflects the extent to which subjectively poor people (by our definition) put vignette 1 above the lowest rung of the ladder. If everyone who is counted as poor by our new definition ($R_i \leq R_i^1$) agrees that vignette 1 is on the lowest rung of the ladder then the two measures will be equal.

Note that there is no obvious basis for making cross-country comparisons using either method of measuring subjective poverty. Here we are interested in comparing the two methods within each country.

From this approach, a simple measure of scale heterogeneity among the poor can be proposed. Consider first the proportion of those who are subjectively poor by our definition and put vignette 1 on the lowest rung. (We ignore differences in the rankings given to vignette 1 among those who assign $R_i^1 > 1$.) Denote this proportion by:

$$P = \frac{\sum_{i=1}^N 1(R_i^1 = 1 | R_i \leq R_i^1)}{\sum_{i=1}^N 1(R_i \leq R_i^1)} \quad (4)$$

where $1(\cdot)$ takes the value 1 if the term in parentheses is true and zero otherwise. A natural measure of scale heterogeneity among the poor (*SHP*) is then the scaled variance:

$$SHP \equiv 4P(1 - P) \in [0,1] \quad (5)$$

At the lower bound of 0 all the poor put vignette 1 on the same rung or none do. At the upper bound of 1 the variance of poor peoples' rankings of vignette 1 is at its maximum, for which half put vignette 1 on the lowest rung.

3.2 *Empirical implementation*

Table 2 provides the distributions of respondents' subjective welfare levels relative to each vignette. The self-assessments provided after hearing the vignettes are probably of greater interest in this context, so we focus the discussion on those results.

We find that 14% of the Tajikistan respondents implicitly felt that they were no better off than the poorest vignette household, which can rarely afford meat, has limited heating and warm clothing during the winter, has poorly clothed children, who are sent to work when reaching secondary-school age. (The full details are given in the Appendix.) By contrast, only 7.5% put themselves on the lowest rung. Thus we find considerable heterogeneity in scales among the poor. Of those who judge their own welfare to be no greater than that of the poorest vignette, over half (56%) put that household above the lowest rung of the welfare ladder. The measure of scale heterogeneity among the poor (using equation 5) in Tajikistan is thus very near the maximum variance at $SHP=0.987$.

For the Guatemala sample, 32% of survey respondents reported themselves as no better off than vignette 1, which lives in an adobe house with one room and no latrine, electricity nor running water, eating beans and tortillas, and is unable to afford meat or eggs. (Recall that the sample is from a poor area.) This is closer to the poverty rate of 25% based solely on the subjective welfare responses. In contrast to Tajikistan, we find that about three-quarters (73%) of the Guatemala respondents who were no better off in their perception than the poorest vignette placed that household on the lowest rung of the welfare ladder. The measure of scale heterogeneity among the poor is 0.796.

In Tanzania 25% of the sample put their own welfare at or below that of a family of three illiterate adults and three children, only one of which is in primary school, living in a mud house with no furniture, with no piped water, no land and engaged in casual agricultural labor. The family has one small meal a day and very rarely eats meat or fish. Correcting for scale heterogeneity using the vignettes thus entails a sizable increase in the subjective poverty rate,

from 17% to 25%. For the Tanzania data, about one-third (36%) of the poor put vignette 1 above rung 1, giving $SHP=0.926$.

The choice of vignette 1 as the “poverty line” for these calculations is natural, but it is of interest also to consider the implications of using vignette 2 instead. Using the raw (un-corrected) data the poverty rates are then 34%, 74%, and 44% respectively, while the corresponding vignette-consistent rates are 61%, 84%, and 79%.

While cross-country comparability is questionable (as already noted), it is at least notable that the ranking of the three countries in terms of poverty is identical, with and without the vignette correction for scale heterogeneity. The Guatemala sample is the poorest, followed by Tanzania, and Tajikistan the least poor. This holds for both vignettes 1 and 2 as the reference.

The two measures of subjective poverty also share similar covariates. Table 3 gives probits for the two measures of subjective poverty—the measure in which one deems the respondent to be poor if she puts herself on the lowest rung and our proposed new measure. Despite the substantial scale heterogeneity we find among the subjectively poor, there is a striking similarity in the two probits. There are only a few cases in which a coefficient is significant in one and not the other, and the sizes of the coefficients are generally similar.

4. Tests for bias in subjective welfare regressions

The subjective poverty measures discussed above naturally ignore rankings among the non-poor, in keeping with the usual focus axiom in poverty measurement. We turn next to regression analysis of the full range of welfare rankings.

First we ask whether vignette responses are correlated with covariates commonly found in subjective welfare regressions in the literature, including objective measures of economic welfare. We assume an ordered probit specification, which has become standard in the literature. However, it should be noted that this specification requires a further restriction on the nature of the heterogeneity problem, as described in the model represented by equations (2) and (3). Specifically, we also need to assume that the heterogeneity takes the form of homogeneous shifts in the scales, such that $\tau_i^k - \tau_i^{k-1}$ is constant across all i for each k . Thus we can define a new, scale-transformed, continuous variable, SW_i^{j*} , by adding an appropriate (individual- and

vignette-specific) constant to SW_i^j such that the scales become constant across respondents when applied to the transformed variable, as required by the ordered probit specification.

Our specification for the determinants of this transformed continuous variable (generating the ordinal categorical responses on each vignette) is as follows:

$$SW_i^{j*} = \beta_j \ln PCE_i + \pi_j X_i + \varepsilon_i^j \quad (j=1,4; i=1,\dots,N) \quad (6)$$

Here PCE denotes per capita expenditure, X is a vector of other household-level variables and ε is a normally distributed error term. The latent continuous variable SW_i^{j*} then generates a discrete response on the scale from 1-6 for each vignette with constant scales.

Table 4 summarizes the OP estimates. There are a number of systematic covariates, although the pseudo R^2 's are low, at approximately 0.02. In two of the three countries the pseudo R^2 is highest for the poorest vignette; in the third (Tanzania) it is roughly the same for the poorest and the least poor and both are higher than for the middle vignettes. There is little sign of a clear pattern in one direction. For vignettes 3 and 4 (but not 1 and 2), we find a positive and statistically significant relationship between $\ln PCE$ and the vignette rankings. Both poor and rich tend to agree that vignettes 1 and 2 are poor, but richer households are more likely to give a high welfare ranking to the better-off households described by vignettes 3 and 4. On the other hand smallholders tend to rate the poorest vignette higher than do other households. Geographic characteristics are more likely to be significant for the vignettes higher on the consumption scale (vignettes 3 and 4).

How much do these effects bias the regressions often found in the literature? Our second test tries to assess the robustness of a standard regression for own-reported subjective welfare, employing widely-used covariates from the literature. Analogously to (6) we assume that:

$$SW_i = \beta \ln PCE_i + \pi X_i + \varepsilon_i \quad (i=1,\dots,N) \quad (7)$$

This is the latent continuous variable for the subjective welfare of respondent i , which generates a discrete response on the scale from 1-6. We refer to the estimated β as the economic gradient in subjective welfare. We do not attempt to make the covariates X identical across the three countries. There is little obvious reason to do so, and reasons for adapting the model to each context. (For example, migration is an important factor in the region of study in Guatemala. And geographic effects are less relevant to our Tanzania data, as they come from just one district.)

To test the robustness of the standard test based on (7), we allow for systematic covariates of the vignettes under the assumption of internal consistency described in Section 3.1, but now relaxing our assumption of homogeneous shifts in the scales. To do so we employ the method of Compound Hierarchical Ordered Probit, proposed by King et al., (2004), dubbed CHOPIT. In the standard model, one postulates the existence of a series of common cut-off points in the *SW* space that generate the observed ordinal responses (as in any OP). Instead, CHOPIT postulates that these thresholds are functions of a vector of observed covariates. The extra information on the vignette responses provides the basis for identification, under the internal consistency assumption—specifically that the thresholds for a respondent’s self-assessed welfare are determined identically as for that respondent’s thresholds in the vignette responses. Following King et al. (2004), the thresholds are assumed to be given by:¹²

$$\tau_i^1 = \gamma^1 \ln PCE_i + \delta^1 X_i \quad (8.1)$$

$$\tau_i^k = \tau_i^{k-1} + \exp(\gamma^k \ln PCE_i + \delta^k X_i) \text{ for } k=2, \dots, K \quad (8.2)$$

The identifying assumption is that the same parameters, γ^k and δ^k , and (hence) the thresholds τ_i^k determine the ordinal responses on the vignettes. Such response consistency is a natural assumption to make. Without the vignettes, identification would only be possible under questionable assumptions about the nonlinearity of the functional forms involved. Thus we are able to model determinants of the thresholds separately to those of the latent continuous variable for subjective welfare.¹³

We begin by testing for an economic gradient in subjective welfare. Table 5 gives the estimated economic gradient (the regression coefficients on log consumption per capita) for various specifications, comparing OP and CHOPIT, but without any controls in the subjective welfare regression. As can be seen from Table 5, the OP and CHOPIT estimates of the coefficients of the equation for subjective welfare turn out to be quite close. Correcting for scale heterogeneity attenuates the economic gradient (lower β) for Tajikistan and Guatemala but

¹² Notice that no error terms appear in the following equations. These are taken to be subsumed in the overall error term ε_i in equation (7).

¹³ We implemented the CHOPIT analysis using the R statistical analysis program, using the programs 'anchors,' 'rgenoud,' and 'Zelig.' Further information and documentation on these packages is available at <http://sekhon.berkeley.edu/rgenoud>, <http://wand.stanford.edu/anchors> and <http://gking.harvard.edu/zelig>. The R code is available on request.

increases it in Tanzania. The post-vignette coefficient estimates are closer than the pre-vignette estimates for Tajikistan and Tanzania but there is little difference for Guatemala.

Table 6 presents the results for an extended specification including various other covariates often found in subjective welfare regressions in the literature, for both OP and CHOPIT. Again we provide both OP and CHOPIT for both pre and post vignettes. And, as found in the previous table, the results are quite similar between the two for each country. (For brevity, we do not comment on the regressions themselves, though they accord reasonably well with our priors based on similar regressions in the literature.) While there are rather few differences between the OP and CHOPIT results, the more notable differences are as follows:

(i) Tajikistan: The significant negative effect in the OP of being a female-headed household is not robust to allowing for scale heterogeneity, although here too there is more similarity in the post-vignette case, with a significant negative effect indicated (comparing the OP and CHOPIT coefficients in Table 6(a)). Nor is the significant negative effect of being Russian robust (whether pre- or post-vignette). Household size has a stronger (positive) effect on subjective welfare when we use CHOPIT. And one of the geographic effects (living in urban Khatlon) becomes much stronger.

(ii) Guatemala: The economic gradient in subjective welfare falls when we adjust for scale heterogeneity (Table 6(b)). A stronger female respondent effect emerges when we correct for bias using CHOPIT, as does the (negative) effect of being single rather than married. The geographic effects also change.

(iii) Tanzania: The economic gradient in subjective welfare rises when we adjust for scale heterogeneity and a much larger (negative) effect of children emerges (Table 6(c)).

Again, we found that a number of covariates were significant predictors of the scales (significant $\hat{\gamma}^k$'s), as revealed by the vignettes although there is little clear pattern. Table 7 summarizes these results and is self-explanatory.

5. Conclusions

There are *a priori* grounds for questioning past applications of survey-based subjective assessments of welfare in measuring and modeling poverty and in calibrating welfare functions. Not only are the interpretations given by households to the scales used in the survey questions likely to vary—casting doubt on the meaning of the derived measures—but there are reasons to

expect these differences to confound inferences about welfare impacts. For example, poorer people may well have more limited horizons in life, stemming from more limited experiences with the extent of the disparities in levels of living in society as a whole. Such differences in knowledge and experience could well translate into a difference in the interpretation given to the scales used in questions on subjective welfare. In particular, poorer people might be expected to use lower thresholds for defining poverty. This would confound efforts to measure poverty and identify welfare effects using subjective data. Some observers have concluded that such data should not be used as dependent variables—casting doubt on a large literature, and warning against future applications.

But is this a serious problem for the many applications of subjective welfare data, including in measuring subjective poverty and estimating welfare effects? The paper has tried to answer that question. An approach to measuring and modeling subjective poverty and welfare has been proposed that takes scale heterogeneity seriously. Our approach relies on carefully designed vignettes on hypothetical households, which were added to household surveys including questions on own-welfare and relevant covariates. Respondents scored the vignettes on the same ladder used to report their own subjective economic welfare.

For measuring subjective poverty, instead of asking what proportion of respondents say that they are on the lowest rung of the welfare ladder, we ask what proportion report that their own welfare is no greater than the poorest vignette. In doing so, subjective poverty measures increase given the presence of scale heterogeneity. The increase is large in two of the three countries. There is considerable scale heterogeneity among the poor. While the overall poverty ranking of the three countries is unaffected, the large change in levels resulting from addressing scale heterogeneity raises concerns over using a simple subjective welfare measure to assess and compare poverty. However, on comparing regression models for our new method of measuring subjective poverty with the past method ignoring scale heterogeneity, we find little difference in the coefficients on the covariates of poverty or in their statistical significance. While there is ample scale diversity, its systematic component does not seriously confound inferences from prevailing methods that ignore this problem.

We do find some significant covariates for vignette responses among a set of regressors commonly used to explain subjective welfare, although the effects defy any simple pattern of bias. To explore further the extent of the overall bias in estimates of the regression coefficients of

subjective welfare on standard covariates, we have compared them with a model that explicitly allows for the heterogeneity in scales. For this purpose, the thresholds were modeled as functions of covariates, assuming consistency between own-welfare scoring and scoring of the vignettes. While some differences are notable, when taken overall, our results suggest quite similar factors influencing subjective welfare when comparing the standard regressions with those augmented to allow for systematic scale heterogeneity. This holds in the data studied for all three countries.

Our findings suggest that scale heterogeneity is a serious concern when using subjective welfare data for making inter-personal comparisons of welfare. The meaning of widely-used subjective poverty measures appears highly questionable. However, more encouragingly, scale heterogeneity is not as great a concern as some observers have claimed for using subjective poverty and welfare measures as dependent variables in situations in which there is no option but to assume constant scales. It seems that, despite scale heterogeneity, one can learn something that is reasonably robust about trade-offs from such data—trade-offs that are otherwise hard to identify.

It should be emphasized that we have deliberately focused here on subjective economic welfare in developing countries. Applications in rich countries or on broader concepts of welfare such as “happiness” or “satisfaction with life” may well entail greater latent heterogeneity in scales, although for happiness it is possibly less obvious that this heterogeneity would be correlated with standard regression covariates. One might try to develop vignettes for rich-country settings or for broader welfare concepts. However, even aside from the difficulties of having long vignettes in surveys, adding many more dimensions into the vignettes will make it less likely that an unambiguous welfare ranking of the vignettes is possible. So it might reasonably be argued that we have picked the “low-lying fruit” of the scale-heterogeneity problem in subjective welfare questions. How best to reach the higher fruit remains as a topic for future research.

Table 1: Pre-vignette and post-vignette subjective welfare rankings

Tajikistan

Pre-vignette	Post-vignette						Total
	1 poorest	2	3	4	5	6 richest	
1 poorest	247	64	23	3	4	0	341
2	72	933	232	53	7	5	1,302
3	34	242	1,735	223	37	4	2,275
4	4	33	112	592	34	1	776
5	2	0	2	11	51	1	67
6 richest	0	0	1	1	4	4	10
Total	359	1,272	2,105	883	137	15	4,771

Note: Pearson $\chi^2(25) = 8.0e+03$ Pr = 0.000; likelihood-ratio $\chi^2(25) = 4.9e+03$ Pr = 0.000; Cramér's V = 0.5806; gamma = 0.8532 ASE = 0.009; Kendall's tau-b = 0.6991 ASE = 0.009.

Guatemala

Pre-vignette	Post-vignette						Total
	1 poorest	2	3	4	5	6 richest	
1 poorest	265	118	17	7	4	1	412
2	37	422	103	7	2	0	571
3	7	50	128	22	3	1	211
4	1	5	4	9	1	0	20
5	0	0	0	2	3	0	5
6 richest	0	0	0	0	1	0	1
Total	310	595	252	47	14	2	1,220

Note: Pearson $\chi^2(25) = 1.1e+03$ Pr = 0.000; likelihood-ratio $\chi^2(25) = 802.2182$ Pr = 0.000; Cramér's V = 0.4287; gamma = 0.8000 ASE = 0.022; Kendall's tau-b = 0.6049 ASE = 0.021

Tanzania

Pre-vignette	Post-vignette						Total
	1 poorest	2	3	4	5	6 richest	
1 poorest	59	31	12	3	1	1	107
2	12	91	52	12	1	0	168
3	6	26	87	26	3	0	138
4	0	1	7	17	4	0	29
5	0	0	2	0	1	2	5
6 richest	0	0	0	0	0	2	2
Total	77	149	150	58	10	5	449

Note: Half-steps included in the survey question have been aggregated into one step (the one below) to facilitate comparisons. Pearson $\chi^2(45) = 551.2519$ Pr = 0.000; likelihood-ratio $\chi^2(45) = 301.5191$ Pr = 0.000; Cramér's V = 0.4955; gamma = 0.6892 ASE = 0.036; Kendall's tau-b = 0.5446 ASE = 0.032.

Table 2: Subjective economic welfare relative to the vignettes

	Tajikistan	Guatemala	Tanzania
Pre-Vignette Self-Assessment			
At or below vignette 1	0.139	0.399	0.346
At or below vignette 2	0.606	0.835	0.792
At or below vignette 3	0.948	0.980	0.984
At or below vignette 4	0.992	0.990	1.000
Post-Vignette Self-Assessment			
At or below vignette 1	0.139	0.319	0.249
At or below vignette 2	0.596	0.787	0.613
At or below vignette 3	0.937	0.971	0.882
At or below vignette 4	0.986	0.987	0.999

Note: The table gives the proportion of respondents in each country who rated their own subjective economic welfare at or below the level they assigned to each vignette (as given in the Appendix).

Table 3(a): Probits for alternative measures of subjective poverty in Tajikistan

	(1)		(2)	
	Probit for placing oneself on step 1		Probit for placing oneself at or below vignette 1	
	Coeff.	s.e.	Coeff.	s.e.
log per capita real consumption	-0.685***	0.082	-0.674***	0.067
<i>Household Head Demographics</i>				
female headed household	0.076	0.087	0.083	0.075
age of hh head	0.003	0.003	0.002	0.003
<i>Ethnicity (Reference: Tajik)</i>				
Uzbek	-0.079	0.082	-0.073	0.068
Russian	0.137	0.228	0.381*	0.219
Other	0.615***	0.194	0.543***	0.178
<i>Education (Reference: No Education)</i>				
Primary	-0.105	0.174	0.021	0.154
Basic	0.024	0.159	0.122	0.144
General Secondary	-0.114	0.157	0.030	0.142
Special Secondary	-0.179	0.174	-0.028	0.155
Technical Secondary	-0.147	0.185	0.058	0.159
Higher Education	-0.290	0.182	-0.146	0.160
Graduate School				
<i>Household Characteristics</i>				
log household size	-0.491***	0.080	-0.561***	0.071
no. of elderly (65+)	-0.031	0.074	0.018	0.061
any migrant(s) in the hhs	-0.017	0.074	0.035	0.062
no. of employed	0.001	0.034	0.017	0.026
Agriculture / Fishing / Forestry	-0.153	0.115	0.038	0.091
Manufacture / Mining	-0.250	0.164	-0.190	0.143
Services	n.a.	n.a.	-0.457**	0.232
Construction	-0.110	0.149	0.061	0.117
Public Administration / Education / Health	-0.080	0.123	-0.106	0.110
Sales and Services	-0.211*	0.120	-0.182*	0.096
Other	0.027	0.170	0.182	0.144
<i>Landholding (Reference: No Land)</i>				
Small Holding	-0.107	0.096	-0.069	0.081
Medium Holding	-0.260**	0.124	-0.210**	0.103
Large Holding	-0.328***	0.125	-0.214**	0.099
<i>Geography</i>				
Sogd Urban	-0.028	0.143	-0.015	0.119
Sogd Rural	0.214	0.137	0.247**	0.116
Khatlon Urban	0.091	0.183	0.441***	0.138
Khatlon Rural	0.473***	0.132	0.405***	0.111
RRP Urban	0.486***	0.165	0.371**	0.147
RRP Rural	0.499***	0.125	0.583***	0.104
Gbao Urban	-0.328	0.344	-0.353	0.263
Gbao Rural	-0.175	0.146	0.101	0.116
Pseudo R ²	0.099		0.087	

Note: *** p<0.01, ** p<0.05, * p<0.1; Col (1): there were no households with heads in services.

Table 3(b): Probits for alternative measures of subjective poverty in Guatemala

	(1)		(2)	
	Probit for placing oneself on step 1		Probit for placing oneself at or below vignette 1	
	Coeff.	s.e.	Coeff.	s.e.
<i>Household Characteristics</i>				
log annual per capita consumption	-0.507***	0.097	-0.426***	0.124
log household size	-0.374***	0.105	-0.384***	0.099
female headed household	-0.463	0.332	-0.498*	0.284
<i>Respondent Characteristics</i>				
respondent is head of household	0.294	0.308	0.521*	0.290
respondent is female	0.270	0.363	0.567*	0.318
age of respondent	-0.004	0.003	-0.002	0.004
year of education of respondent	-0.147***	0.023	-0.074**	0.035
respondent is employed	0.113	0.176	0.108	0.163
single / never married	-0.508**	0.237	-0.684***	0.254
union	-0.744***	0.261	-0.845***	0.265
separated/divorced	0.511	0.433	0.059	0.402
widowed	-0.199	0.312	-0.476	0.321
Municipality (reference: Cuilco)				
San Gaspar Ixchil	-0.204	0.167	-0.118	0.127
Santa Ana Huista	-0.783***	0.181	-0.602***	0.118
Jacaltenango	0.169	0.171	0.144	0.223
<i>Migration characteristics</i>				
any migrant LA (excluding respondent)	0.212	0.212	0.386*	0.207
any migrant US (excluding respondent)	-0.238	0.229	-0.089	0.203
migrated internally	-0.026	0.187	-0.210	0.195
migrated in Latin America	0.466***	0.160	0.398***	0.148
migrated to USA	-0.256	0.238	0.078	0.197
Pseudo R ²	0.182		0.130	

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 3(c): Probits for alternative measures of subjective poverty in Tanzania

	(1)		(2)	
	Probit for placing oneself on step 1		Probit for placing oneself at or below vignette 1	
	Coeff.	s.e.	Coeff.	s.e.
<i>Household Characteristics</i>				
log per capita consumption	-0.431***	0.126	-0.556***	0.117
log household size	-0.038	0.244	-0.122	0.285
share of women	0.299	0.514	-0.261	0.454
share of children under age 6	-0.327	0.961	-1.180	1.026
share of children between 6 and 15	0.113	0.568	-0.154	0.621
share of members 55 and older	-0.914	0.637	-0.468	0.613
log land size (in acres)	-0.397***	0.101	-0.321***	0.093
<i>Household Head Characteristics</i>				
household head female	1.087***	0.321	0.938***	0.288
household head age	0.013	0.011	0.012	0.010
household head years of education	0.007	0.071	0.119*	0.069
<i>Respondent Characteristics</i>				
respondent female	1.121***	0.252	0.981***	0.225
respondent age	0.007	0.012	-0.003	0.009
respondent years of education	-0.061	0.077	-0.180**	0.078
Pseudo R ²	0.185		0.186	

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 4: Significant predictors of how households rank the four vignettes

	Vignette			
	1 (poorest)	2	3	4 (richest)
Tajikistan				
Household size (+)	Special secondary schooling (+)	Log Consumption p.c . (+)	Log Consumption p.c. (+)	
Basic education (-)	Number of employed (+)	Uzbek (+)	Uzbek (+)	
Services sector occupation (+)	Small holding (-)	Primary schooling (+)	Number of employed (+)	
Small holding (+)	Sogd (+)	General secondary (+)	Agriculture sector (-)	
Khatlon urban (-)	Khatlon (+)	Number of employed (+)	Small and med. holding (-)	
Gbao rural (-)	Gbao urban (-)	Small holding (-)	Sogd (-)	
	Gbao rural (+)	Sogd rural (-)	Khatlon urban (+)	
		Khatlon urban (+)	Khatlon rural (-)	
		RRP rural (-);Gbao (+)	RRP (-);Gbao rural (+)	
Pseudo R ² 0.022	0.018	0.016	0.019	
Guatemala				
Female respondent (+)	HH owns land (+)	HH owns land (+)	HH owns land (+)	
Marital Status Union (-)	Female Headed Household (-)	Marital Status Union (-)	Years of Education (-)	
Marital Status Married (-)	Respondent is HH head (+)	Divorce/Separated (-)	Household migrant to Latin America (-)	
Divorce/Separated (-)	Female respondent (+)	Santa Ana Huista (-)	Cuilco (-)	
Widowed (-)	Marital Status Union (-)	Jacaltenango (-)	Santa Ana Huista (-)	
HH migrant to Latin America (-)	Marital Status Married (-)		Jacaltenango (-)	
Internal migrant (-)	Divorce/Separated (-)			
Migrant to USA (+)	Widowed (-)			
Santa Ana Huista (-)	Migrant to USA (+)			
	Cuilco (+); Jacaltenango (-)			
Pseudo R ² 0.05	0.03	0.025	0.045	

Tanzania

Log Consumption p.c. (-)	Log Consumption p.c. (-)	Log Consumption p.c. (-)	Log Consumption p.c. (-)
Share of Women (-)	Share of Children under 6 (+)	Years of Education (-)	Respondent Female (-)
Log Land Size (+)	Log Land Size (+)	Share of Children 6-15 (-)	Share of Children 6-15 (-)
Pseudo R ² 0.0266	0.0164	0.0174	0.0301

Note: The table gives significant covariates at the 10% level or better. See the Appendix for details on the vignettes.

Table 5: Testing for an economic gradient in subjective welfare

Regression coefficients on log expenditure per person	Ordered Probit		CHOPIT	
	Coeff.	s.e.	Coeff.	s.e.
Tajikistan				
<i>Pre-vignette</i>	0.630***	0.037	0.580***	0.033
<i>Post-vignette</i>	0.560***	0.037	0.513***	0.032
Guatemala				
<i>Pre-vignette</i>	0.429***	0.087	0.390***	0.055
<i>Post-vignette</i>	0.444***	0.065	0.389***	0.053
Tanzania				
<i>Pre-vignette</i>	0.419***	0.117	0.504***	0.085
<i>Post-vignette</i>	0.529***	0.105	0.560***	0.082

Table 6(a): Subjective welfare regressions for Tajikistan

	(1)		(2)		(3)		(4)	
	Pre-vignette				Post-vignette			
	Ordered probit		CHOPIT		Ordered probit		CHOPIT	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
log per capita real consumption	0.729***	0.042	0.731***	0.039	0.650***	0.041	0.650***	0.038
<i>Household Head Demographics</i>								
female headed household	-0.155***	0.052	-0.069	0.051	-0.176***	0.052	-0.110**	0.049
age of hh head	-0.001	0.002	0.001	0.002	-0.002	0.002	0.000	0.0018
<i>Ethnicity (Reference: Tajik)</i>								
Uzbek	0.020	0.046	0.031	0.048	0.064	0.045	0.056	0.0468
Russian	-0.334**	0.158	-0.148	0.159	-0.410**	0.161	-0.186	0.1544
Other	-0.690***	0.158	-0.481***	0.137	-0.647***	0.158	-0.396***	0.132
<i>Education (Reference: No Education)</i>								
Primary	0.065	0.117	-0.138	0.115	-0.015	0.113	-0.259**	0.111
Basic	-0.134	0.109	-0.150	0.108	-0.193*	0.108	-0.271***	0.105
General Secondary	0.006	0.106	-0.037	0.105	-0.098	0.104	-0.184*	0.101
Special Secondary	0.120	0.113	0.084	0.112	-0.032	0.111	-0.128	0.108
Technical Secondary	0.077	0.116	0.056	0.114	-0.079	0.113	-0.125	0.1103
Higher Education	0.283**	0.114	0.262**	0.112	0.180	0.111	0.103	0.1078
Graduate School	0.542	0.375	0.784*	0.443	0.516	0.400	0.601	0.4281
<i>Household Characteristics</i>								
log household size	0.072***	0.009	0.492***	0.048	0.064***	0.009	0.457***	0.047
no. of elderly (65+)	-0.001	0.043	-0.031	0.041	-0.028	0.042	-0.036	0.0397
any migrant(s) in the hhs	0.045	0.043	-0.004	0.042	0.068	0.043	0.012	0.0405
<i>Employment Characteristics of Household Head (Reference for Occupation: Unemployed)</i>								
No. of employed	0.036**	0.017	0.014	0.017	0.010	0.018	-0.005	0.0168
Agriculture / Fishing / Forestry	-0.002	0.063	0.005	0.062	0.062	0.062	0.087	0.0598
Manufacture / Mining	-0.033	0.094	0.013	0.095	0.053	0.092	0.096	0.092
Services	0.351***	0.130	0.355**	0.147	0.442***	0.119	0.367**	0.143
Construction	-0.072	0.080	-0.085	0.080	-0.032	0.077	-0.066	0.0778
Public Administration / Education / Health	0.175**	0.069	0.134**	0.066	0.092	0.066	0.089	0.0643
Sales and Services	0.303***	0.064	0.249***	0.061	0.236***	0.062	0.200***	0.059
Other	-0.167	0.107	-0.041	0.106	-0.142	0.107	-0.033	0.1029

<i>Landholding (Reference: No Land)</i>								
Small Holding	-0.068	0.058	0.004	0.057	-0.049	0.056	0.022	0.0549
Medium Holding	-0.110	0.069	-0.040	0.070	0.043	0.069	0.101	0.0679
Large Holding	0.014	0.070	0.071	0.068	0.076	0.067	0.103	0.0655
<i>Geography</i>								
Sogd Urban	-0.011	0.076	-0.055	0.081	-0.015	0.072	-0.059	0.0785
Sogd Rural	-0.109	0.079	-0.071	0.080	-0.091	0.078	-0.057	0.0773
Khatlon Urban	-0.186*	0.095	-0.618***	0.096	-0.093	0.089	-0.499***	0.093
Khatlon Rural	-0.511***	0.074	-0.593***	0.076	-0.472***	0.073	-0.549***	0.073
RRP Urban	-0.161	0.108	-0.143	0.107	-0.287***	0.107	-0.256**	0.105
RRP Rural	-0.265***	0.075	-0.300***	0.073	-0.245***	0.076	-0.288***	0.071
Gbao Urban	0.340***	0.107	0.304**	0.125	0.567***	0.115	0.432***	0.121
Gbao Rural	-0.067	0.078	-0.143*	0.081	0.029	0.078	-0.065	0.0786

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 6(b): Subjective welfare regressions for Guatemala

	(1)		(2)		(3)		(4)	
	Pre-vignette				Post-vignette			
	Ordered probit		CHOPIT		Ordered probit		CHOPIT	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Household Characteristics</i>								
log annual per capita consumption	0.449***	0.080	0.351***	0.066	0.507***	0.076	0.402***	0.064
log household size	0.223**	0.098	0.173*	0.093	0.267**	0.105	0.309***	0.090
female headed household	0.464**	0.228	0.561**	0.228	0.467**	0.229	0.489**	0.221
<i>Respondent Characteristics</i>								
respondent is head of household	-0.417*	0.222	-0.515**	0.223	-0.423*	0.223	-0.399*	0.216
respondent is female	-0.275	0.252	-0.440*	0.241	-0.375	0.255	-0.435*	0.234
age of respondent	0.002	0.003	0.001	0.003	0.002	0.003	0.001	0.003
year of education of respondent	0.073***	0.020	0.091***	0.019	0.073***	0.019	0.093***	0.018
respondent is employed	0.195	0.136	0.147	0.124	-0.068	0.127	-0.0322	0.122
<i>Respondent Marital Status (reference: married)</i>								
single / never married	-0.555**	0.227	-1.063***	0.244	-0.531**	0.241	-0.951***	0.233
union	-0.175**	0.083	-0.171**	0.086	-0.175**	0.079	-0.170**	0.084
separated/divorced	-0.930**	0.411	-0.474	0.306	-1.035***	0.318	-0.808***	0.310
widowed	-0.550***	0.174	-0.481***	0.171	-0.418**	0.182	-0.395**	0.163
<i>Municipality (reference: Cuilco)</i>								
San Gaspar Ixchil	0.067	0.106	0.653***	0.109	-0.060	0.108	0.477***	0.106
Santa Ana Huista	0.484***	0.108	0.131	0.105	0.236**	0.094	0.061	0.101
Jacaltenango	0.110	0.157	0.599***	0.147	-0.005	0.157	0.550***	0.143
<i>Migration characteristics</i>								
any migrant LA (excluding respondent)	-0.163	0.140	-0.156	0.118	-0.013	0.130	-0.155	0.114
any migrant US (excluding respondent)	0.500***	0.131	0.595***	0.132	0.313**	0.145	0.556***	0.128

migrated internally	-0.181	0.130	0.054	0.140	-0.190	0.124	0.094	0.136
migrated in Latin America	-0.360**	0.143	-0.227*	0.117	-0.403***	0.132	-0.193*	0.113
migrated to USA	0.003	0.150	-0.051	0.149	0.120	0.151	0.169	0.144

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 6(c): Subjective welfare regressions for Tanzania

	(1)		(2)		(3)		(4)	
	Pre-vignette				Post-vignette			
	Ordered probit		CHOPIT		Ordered probit		CHOPIT	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Household Characteristics</i>								
log per capita consumption	0.301***	0.101	0.421***	0.099	0.386***	0.110	0.498***	0.095
log household size	0.432*	0.234	0.385*	0.201	0.207	0.206	0.507***	0.192
share of women	-0.231	0.338	0.361	0.306	0.232	0.307	0.384	0.288
share of children under age 6	-0.175	0.582	-0.558	0.512	-0.182	0.590	-1.030**	0.484
share of children between 6 and 15	-0.412	0.605	-0.186	0.437	0.138	0.472	-0.603	0.412
share of members 55 and older	0.413	0.553	0.000	0.475	0.565	0.548	0.199	0.452
log land size (in acres)	0.193**	0.087	0.161**	0.075	0.158**	0.079	0.169**	0.071
<i>Household Head Characteristics</i>								
household head female	-0.324	0.262	-0.369	0.237	-0.671***	0.249	-0.568**	0.226
household head age	-0.009	0.008	-0.003	0.009	-0.012*	0.007	-0.011	0.009
household head years of education	0.046	0.050	0.064*	0.038	0.046	0.038	0.041	0.037
<i>Respondent Characteristics</i>								
respondent female	-0.399**	0.166	-0.314*	0.180	-0.538***	0.170	-0.456***	0.173
respondent age	0.009	0.008	0.005	0.009	0.008	0.007	0.004	0.008
respondent years of education	0.052	0.055	0.026	0.038	0.044	0.036	0.034	0.036

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 7: Significant predictors of scale heterogeneity

	γ^1	γ^2	γ^3	γ^4	γ^5
Tajikistan Significant covariates at the 10% level or better	pre-vignette log PCE (-)	Log PCE (+)	Uzbek (-)	Employed (+) "Other" sector employment (-)	Russian (-) Secondary education (+) Public administration employment (-)
	Secondary education (-) Graduate School /Aspitantura (+)	Primary education (+) Household size (+) Services sector occupation (+)	Sogd rural (-)	Sogd urban (+)	Sogd rural (-)
	Household size (-) Services sector occupation (-) Sogd urban (-) RRP rural (-)	Medium land holding (+) Large land holding (+) Khatlon urban (-)		Sogd rural (+) Khatlon urban (+) Khatlon rural (-) RRP rural (+)	
	Tajikistan post-vignette Gamma cut 1	Gamma cut 2	Gamma cut 3	Gamma cut 4	Gamma cut 5
	Secondary education (-) Household size (-) Service sector employment (-)	Household size (+) Employed (-) Service sector employment (+)	Uzbek (-) Sogd urban (+) Sogd rural (+)	Household size (+) Employed (-) Public administration employment (+) "Other" sector employment (-) Sogd urban (+) Sogd rural (+) Khatlon urban (+) Khatlon rural (+) RRP rural (+) Gbao rural (+)	Log PCE (+) Russian (-) Secondary education (+) Public Administration employment (-) Sogd rural (+)
	Sogd urban (-) Khatlon urban (-) Gbao urban (+)	Small land holding (+) Medium land holding (+) Large land holding (+) Khatlon urban (-) Khatlon rural (-) Gbao rural (-)	Gbao urban (-)		
	Guatemala pre-vignette household size (+)	household size (+)	respondent divorced (-)	log PCE (-)	Household size (+)
respondent female (-)	household member migrated previously to Latin America (excluding respondent) (+)	respondent widowed (-)	Santa Ana Huista (+)	Respondent age (-)	
respondent in union (+)	Cuilco (+)	respondent previously migrated to Latin America (-)		Respondent is employed (-)	

	respondent married (+) respondent divorced (+) respondent widowed (+) respondent migrated to US (-) Cuilco (-) Santa Ana Huista (-) Jacaltenango (-)	Santa Ana Huista (+)	Santa Ana Huista (+)		Cuilco (-) Jacaltenango (-)
Guatemala	post-vignette respondent is female (-)	household member migrated previously to Latin America (excluding respondent) (-)	respondent in union (-)	log PCE (-)	household size (+)
	respondent in union (+) married (+) divorced (+) widowed (+) previously migrated within Guatemala (+) previously migrated to US (-)	respondent years of education (+) respondent in union (+) married (+) divorced (+) Cuilco (+) Santa Ana Huista (+)	respondent married (-) divorced (-) previously migrated to Latin America (+) Santa Ana Huista (+)	household size (-) Santa Ana Huista (+)	respondent age (-) Cuilco (-) Jacaltenango (-)
	Cuilco (-) Santa Ana Huista (-) Jacaltenango (+)				
Tanzania	pre-vignette share of women in h'hold (+)				share of children 6 - 15 (-)
			log PCE (+) share of women (-) share of young children in household (+) share of children 6 - 15 (+)		
Tanzania	post-vignette respondent female (+) share of women in h'hold (+) land size (-)	land size (+)	log PCE (+)		

Note: The table gives the significant (5% level) γ^k coefficients from the CHOPIT regressions

Appendix: The vignettes

	Tajikistan	Guatemala	Tanzania
1	Family A can only afford to eat meat on very special occasions. During the winter months, they are able to partially heat only one room of their home. They cannot afford for children to complete their secondary education because the children must work to help support the family. When the children are able to attend school, they must go in old clothing and worn shoes. There is not enough warm clothing for the family during cold months. The family does not own any farmland, only their household vegetable plot.	Family Castillo lives in an adobe house with one room and no latrine. The house does not have electricity or running water. The family eats beans and tortillas, but is never able to afford meat, eggs.	Joseph's/Josephine's family has 6 people – 3 adults and 3 children – living in a mud house with the river as the main source of water. One of the children is in primary school. None of the adults are literate. The family has no land and supports itself by engaging in casual agricultural labor for a large landowner. They have one small meal a day and very rarely eat <i>matooke</i> , meat or fish. The family has no furniture and sleeps on the floor.
2	Family B can afford to eat meat only once or twice a week. During winter months, they can heat several rooms, but not the entire house. They cannot afford for all their children to complete secondary education. Their clothing is sufficiently warm, but they own only simple garments. In addition to their household vegetable plot, they own a small plot of poor quality farmland that is distant from their home.	Family Gomez lives in an adobe house with two rooms and a latrine. The house has electricity but no running water. The family owns a bicycle and small battery-powered radio. They eat mainly beans, eggs, tortilla, rice and corn.	Edward's/Esther's family has 6 people – 3 adults and 3 children – living in a mud house with the river as the main source of water. One of the children is in primary school. None of the adults are literate. The family has a one acre banana plantation. The adult male does some casual labor in construction in town. The family eats two small meals a day, and is able to occasionally eat meat or <i>dagaa</i> . The family has three old mattresses, a bench for guests and a few chickens.
3	Family C can afford to eat meat every day. During the winter months, generally they are able to keep their home warm. They can afford for all their children to complete secondary education. They have sufficient clothing to keep warm in the winter. Their everyday clothing is simple, but they also have some fancy items for special occasions. In addition to their household vegetable plot, they have a larger plot of good quality farmland, not too distant from their home.	Family Hernandez lives in a block house with a-iron sheet roof. The house has two rooms, a latrine, running water and electricity. The family owns a used motorcycle, two bicycles, a TV, refrigerator and small stereo. They are able to eat meat or chicken at least twice a week.	Godi's/Rose's family has 6 people - 3 adults and 3 children – living in an un-cemented brick house with access to the community water stand. Two of the children are in primary school. None of the adults are literate. The family has a 2.5 acre banana plantation. Two adults do some part time casual labor in town. The family eats three meals a day and is often able to eat meat and fish. The family has three beds with thin mattresses, one bench, a bicycle and some chicken.
4	Family D can afford to eat whichever foods they would like, including sweets and imported food. During the winter months, they have no problems with heating and are able to keep their entire house warm. They can afford for all of their children to complete their education, and then to continue at a local university. They are able to afford a variety of fancy traditional clothes and also imported brand clothing. The family owns property, including a good car. The family also has a large farm and acts as landlord to others in their area.	Family Martinez lives in a house with plastered-block walls, a terrace, electricity, and running water. The house has three rooms and a flush bathroom. They have a small store in the community. The family has a truck, TV, refrigerator, stereo and washing machine. They are able to eat meat, chicken or fish daily.	Medard's/Mary's family has 6 people – 3 adults and 3 children – living in a brick and cement house with a tap in the compound. All of the children are in primary school. One adult male is literate. He travels two months of the year to the regional and national capital to engage in trading. The family has a 4 acre banana plantation. The family eats three meals a day which usually include meat or fish. The family has three beds with mattresses, cushioned chairs, a table, radio, clock, second hand motorcycle, two cows and numerous chickens.

References

- Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., O'Donnell, O. 2008. Does Reporting Heterogeneity Bias the Measurement of Health Disparities? *Health Economics*. 17(3), 351-375.
- Beegle, K., Himelein, K. , Ravallion, M. 2012. Frame-of-Reference Bias in Subjective Welfare. *Journal of Economic Behavior and Organization*. 81, 556-570.
- Bertrand, M., Mullainathan, S. 2001. Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review* 91(2), 67-72.
- Bishop, J., Luo, F., Pan, X. 2006. Economic Transition and Subjective Poverty in Urban China. *Review of Income and Wealth*. 52(4), 625-641.
- Carletto, G. and Zezza, A. 2006. Being Poor, Feeling Poorer: Combining Objective and Subjective Measures of Welfare In Albania. *The Journal of Development Studies*. 42(5), 739-760.
- Castilla, C. 2010. Subjective Poverty and Reference-Dependence Income over Time, Aspirations and Reference Groups. UNU-WIDER Working Paper 2010/76.
- Clark, A., Oswald, A. 1994. Unhappiness and Unemployment. *Economic Journal*. 104 (424), 648-59.
- _____ and _____ 1996. Satisfaction and Comparison Income. *Journal of Public Economics* 61, 359-381.
- Conti, G., Pudney, S.. 2011. Survey Design and the Analysis of Satisfaction. *Review of Economics and Statistics*. 93(3), 1087-1093.
- Diener, E., Suh, E., Lucas, R., Smith, H. 1999. Subjective Well-Being: Three Decades of Progress. *Psychological Bulletin*. 125, 276-302.
- Di Tella, R. MacCulloch, R. 2006. Some Uses of Happiness Data in Economics. *Journal of Economic Perspectives*. 20(1), 25-46.
- Dolan, Paul, Tessa Peasgood and Mathew White. 2008. Do we Really Know What Makes us Happy? A Review of the Economic Literature on the Factors Associated with Subjective Well-Being. *Journal of Economic Psychology* 29, 94-122.
- Easterlin, R. 1995. Will Raising the Incomes of all Increase the Happiness of all? *Journal of Economic Behavior and Organization*. 27, 35-47.
- Fafchamps, M., Shipli, F. 2009. Isolation and Subjective Welfare: Evidence from South Asia. *Economic Development and Cultural Change*. 57(4), 641-83.

- Ferrer-i-Carbonell, A. 2005. Income and Well-Being: An Empirical Analysis of the Comparison Income Effect. *Journal of Public Economics*. 89, 997-1019.
- Frank, R. H. 1997. The Frame of Reference as a Public Good. *Economic Journal*. 107, 1832-1847.
- Frey, B., Stutzer, A. 2002. What Can Economists Learn from Happiness Research? *Journal of Economic Literature*. XL, 402-435.
- Furnham, A., Argyle, M. 1998. *The Psychology of Money*. London: Routledge.
- Graham, C., Felton, A. 2006. Inequality and Happiness: Insights from Latin America. *Journal of Economic Inequality* 4(1), 107-122.
- Graham, C., Pettinato, S. 2002. Frustrated Achievers: Winners, Losers, and Subjective Well-being in New Market Economies, *Journal of Development Studies* 38(4), 100-140.
- Herrera, J., Razafindrakoto, M., Roubaud, F. 2006. The Determinants of Subjective Poverty: A Comparative Analysis in Madagascar and Peru. Working Papers DT/2006/01, DIAL.
- Hopkins, D., King, G. 2010. Improving Anchoring Vignettes Designing Surveys to Correct for Interpersonal Incomparability. *Public Opinion Quarterly*. 1-22.
- Kapteyn, A., Kooreman, P., Willemse, R. 1998. Some Methodological Issues in the Implementation of Subjective Poverty Definitions. *Journal of Human Resources*. 23(2), 222-242.
- Kapteyn, A., Smith, J., Van Soest, A. 2008. Comparing Life Satisfaction. Working Paper WR-623, Rand Cooperation.
- King, G., Wand, J. 2007. Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes. *Political Analysis*. 15(1), 46-66.
- King, G., Murray, C., Salomon, J., Tandon, A. 2004. Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review*. 98(1), 191-207.
- Kingdon, G., Knight, J. 2006. Subjective Well-Being Poverty vs. Income Poverty and Capabilities Poverty? *Journal of Development Studies*. 42(7), 1199-1224.
- Kingdon, G., Knight, J. 2007. Community Comparisons and Subjective Well-being in a Divided Society. *Journal of Economic Behavior and Organization*. 64, 69-90.
- Knight, J., Gunatilaka, R. 2010. The Rural-Urban Divide in China: Income but not Happiness? *Journal of Development Studies*. 46(3), 506-534.

_____ and _____, 2012, Income, Aspirations and the Hedonic Treadmill in a Poor Society, *Journal of Economic Behavior and Organization*, forthcoming.

Kristensen, N., Johansson, E. 2008. New Evidence on Cross-Country Differences in Job Satisfaction Using Vignettes. *Labor Economics*. 15(1), 96-117.

Lokshin, M., Ravallion, M. 2008. Testing for an Economic Gradient in Health Status Using Subjective Data. *Health Economics*. 17, 1237-1259.

Luttmer, E. 2005. Neighbors as Negatives: Relative Earnings and Well-Being. *Quarterly Journal of Economics*. 120(3), 963-1002.

Mangahas, M. 1995. Self-rated Poverty in the Philippines. *International Journal of Public Opinion Research*. 7(1), 40-55.

McBride, M. 2001. Relative Income Effects on Subjective Well-Being in the Cross-Section. *Journal of Economic Behavior and Organization*. 45, 251-278.

Oswald, A. 1997. Happiness and Economic Performance. *Economic Journal*. 107, 1815-1831.

Posel, D., Rogan, M. 2013. Measured as Poor *Versus* Feeling Poor: Comparing Objective and Subjective Poverty Rates in South Africa. Paper prepared for the UNU-WIDER Conference Helsinki, Finland.

Pradhan, M., Ravallion, M. 2000. Measuring Poverty Using Qualitative Perceptions of Consumption Adequacy. *Review of Economics and Statistics*. 82(3), 462-471.

Ravallion, M. 2008. On the Welfarist Rationale for Relative Poverty Lines, in Kaushik Basu and Ravi Kanbur (eds) *Social Welfare, Moral Philosophy and Development: Essays in Honour of Amartya Sen's Seventy Fifth Birthday*, Oxford: Oxford University Press.

_____. 2012a, Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty. Policy Research Working Paper 5968, World Bank, Washington DC. (Forthcoming in *Happiness and Economic Growth: Lessons from Developing Countries*, edited by Andrew Clark and Claudia Senik, Oxford University Press.)

_____. 2012b. Poverty Lines Across the World. In *The Oxford Handbook of the Economics of Poverty*, edited by Philip N. Jefferson, Oxford: Oxford University Press.

Ravallion, M., Lokshin, M. 2001. Identifying Welfare Effects from Subjective Questions. *Economica* 68(271): 335-357

_____, _____. 2002. Self-rated Economic Welfare in Russia. *European Economic Review* 46(8): 1453-1473.

_____, _____. 2010. Who Cares About Relative Deprivation? *Journal of Economic Behavior and Organization*. 73(2), 171-185.

Shaffer, P. 2013. Ten Years of “Q-Squared”: Implications for Understanding and Explaining Poverty. *World Development*. 45, 269-285.

Senik, C. 2004. When Information Dominates Comparison: Learning from Russian Subjective Panel Data. *Journal of Public Economics*. 88, 2099-2123.

van de Stadt, H., Kaptyn, A., van de Geer, S. 1985. The Relativity of Utility: Evidence from Panel Data. *Review of Economics and Statistics*. 67, 179-187.

Van Praag, B. 1968. *Individual Welfare Functions and Consumer Behavior*, Amsterdam: North-Holland.

Winkelmann, L., Winkelmann, R. 1998. Why Are the Unemployed So Unhappy: Evidence from Panel Data. *Economica*. 65, 1-15.

World Bank. 2008. Basic Information Document: 2007 Tajikistan Living Standards Measurement Study. Mimeo, World Bank, Washington, DC.

Young Lives Project. 2009. Duy’s story: A profile from Young Lives in Vietnam, Young Lives Project, Department of International Development, University of Oxford, Oxford, England.