Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford

OPHI WORKING PAPER NO. 50

Measurement Errors and Multidimensional Poverty

Cesar Calvo[1] and Fernando Fernandez[2]

January 2012

Abstract

Under quite likely conditions, data measurement errors can cause an (upward) bias in unidimensional poverty estimates and thus mislead both conceptual analyses and policy implications. In the case of multidimensional poverty, we find that by proposing a dual cut-off strategy, the Alkire-Foster method will typically attenuate this bias. With data from a 2010 Living Standard Measurement Survey (LSMS) from Peru, we find empirical evidence in support of this virtue of the dual cut-off strategy.

Keywords: Measurement error, poverty measurement, multidimensional poverty, poverty indices

JEL classification: I32, O15

1 Universidad de Piura, Peru, cesar.calvo@udep.pe

2 Universidad de Piura, Peru, fernando.fernandez@udep.pe

This study has been prepared within the OPHI theme on multidimensional measurement.

OPHI gratefully acknowledges support from UK Economic and Social Research Council (ESRC)/(DFID) Joint Scheme, Robertson Foundation, UNICEF N'Djamena Chad Country Office, United Nations Development Programme (UNDP) Human Development Report Office, Georg-August-Universität Göttingen, UK Department of International Development (DFID), International Food Policy Research Institute (IFPRI), Ministry of Planning Chile, Vanderbilt University Medical Centre, Doris Oliver Foundation, National UNDP and UNICEF offices and private benefactors. International Development Research Council (IDRC) of Canada, Canadian International Development Agency (CIDA), and AusAID are also recognised for their past support.

ISSN 2040-8188 ISBN 978-1-907194-36-8

Acknowledgements

We are grateful to participants at the 2011 OPHI Workshop on Multidimensional Poverty Dynamics for their comments. Errors are our own.

The Oxford Poverty and Human Development Initiative (OPHI) is a research centre within the Oxford Department of International Development, Queen Elizabeth House, at the University of Oxford. Led by Sabina Alkire, OPHI aspires to build and advance a more systematic methodological and economic framework for reducing multidimensional poverty, grounded in people's experiences and values.

Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford
3 Mansfield Road, Oxford OX1 3TB, UK
Tel. +44 (0)1865 271915 Fax +44 (0)1865 281801
ophi@qeh.ox.ac.uk http://ophi.qeh.ox.ac.uk/

1 Introduction

Even though unidimensional poverty remains the bread-and-butter of economic development literature, the need for multidimensional analyses is now widely recognised. Both conceptual works (Alkire, 2002) and proposals for empirical measurement (Atkinson and Bourguignon, 1982; Bourguignon and Chakravarty, 2003; Deutsch and Silber, 2005) have cogently drawn attention towards wellbeing dimensions such as health, culture, leisure, and social life, *inter alia*. In particular, multidimensional measures have raised and addressed several questions on which dimensions should be considered, how they relate to each other, and whether their joint shortfalls should be treated differently from the simple sum of independent, unidimensional shortfalls – see Ravallion (2011) for a review. Among these measures, the proposal by Alkire and Foster (2011) has gained prominence, both for its conceptual underpinnings and for its empirical tractability. It is characterised chiefly by its dual cut-off strategy – failure to reach the poverty line in any dimension is only relevant if such failure is observed in at least a given number of other dimensions.

In this paper, we focus on the impact of measurement errors on this multidimensional measure, P^{AF}. In the unidimensional case, both intuition and formal analysis (Chesher and Schluter, 2002) warn that poverty estimates might be biased in the presence of measurement errors, even if these errors are drawn from symmetrical, zero-mean distributions. We explore whether this caveat carries over to the multidimensional case, and we find that by virtue of the dual cut-off strategy of P^{AF}, biases induced by measurement errors can be presumed to lessen.

In section 2, we discuss the impact of measurement errors on unidimensional poverty estimates. We also introduce the notation and the framework we use as we turn to multidimensional poverty in section 3. Section 4 presents an empirical exercise with Peruvian data in an attempt to spot traces of (unobserved) measurement errors in the data and their impact on observed poverty estimates, both uni- and multidimensional. Section 5 concludes.

2 Unidimensional poverty and measurement error

Poverty measures have been and will continue to be a common tool both for analysis and for policy guidance. As empirical exercises and policy reports have proliferated over time, it is surprising that more attention has not been paid to the consequences of likely measurement errors in the datasets at hand – for instance, the results in Chesher and Schluter (2002) (described below) are rarely referenced, even though they warn unidimensional poverty (as measured by FGT1 and FGT2) will be overrated if consumption is observed with error.

To begin, note a poor individual may be mistaken as non-poor due to measurement error, while someone else above the poverty line may be wrongly seen as deprived. These misjudgements are akin to type-I and type-II errors in statistical inference, respectively.[1] Likewise, the depth of poverty of a poor individual may be overrated (type II, slightly abusing the term) or underrated (type I).

Importantly, these two errors have asymmetric consequences – even if measurement errors are drawn from a symmetrical, zero-mean distribution. They are thus likely to bias poverty measures. The rationale can be put as follows: the censoring of wellbeing outcomes at a poverty line translates into a parallel censoring of measurement error realisations at a certain threshold. We next formalise this intuition.

Let y stand for an outcome measuring wellbeing in any particular dimension, and $\tilde{y}$ for its observed value, including an additive, classical measurement error ε, such that $\tilde{y} = y + \varepsilon$. This error is drawn from a distribution with c.d.f. $\Phi_{(\varepsilon)}$ and density function $\phi_{(\varepsilon)}$, such that $\mathrm{E}[\varepsilon] = \int_{-\infty}^{+\infty} \varepsilon \phi_{(\varepsilon)} \mathrm{d}\varepsilon = 0$.

Absolute poverty is determined by a shortfall below an absolute poverty line z.[2] Thus, it effectively depends on a censored wellbeing outcome y^*:

$$y^* = \mathrm{Min}[y, z] = y + \mathrm{Min}[0, z - y]$$

A poverty measure feeding on data fraught with measurement error will thus

[1] In the analysis of poverty-alleviation programmes, type I and II errors are common terms: both undercoverage of the poor (type I) and leakage of resources into non-poor groups (type II) are sources of inefficiency.

[2] Throughout the paper, by ‘poverty’ we mean *absolute* poverty. Relative poverty lines are not addressed here.

depend on $\tilde{y}^*$:

$$\begin{aligned}\tilde{y}^* &= \text{Min}[\tilde{y}, z]\\ &= \text{Min}[y+\varepsilon, z]\\ &= y + \text{Min}[\varepsilon, z-y]\end{aligned}$$

Hence,

$$\text{E}[\tilde{y}^*] = y + \int_{-\infty}^{z-y} \varepsilon\phi_{(\varepsilon)}\text{d}\varepsilon + (z-y)\left[1-\Phi_{(z-y)}\right]$$

We next ascertain whether observed, censored wellbeing outcomes ($\tilde{y}^*$) can be expected to align with unobserved, true censored outcomes (y^*). We take the cases of non-poor ($z-y<0$) and poor ($z-y>0$) individuals in turn.

2.1 The case of non-poor [NP] individuals: $z-y<0$

If the individual truly is non-poor, then measurement errors can only distort poverty estimates by wrongly inducing the observer to believe the individual lies below the poverty line (type II). The distribution function $\Phi_{(\varepsilon)}$ determines whether this distortion is possible, as well as how bad observed poverty might be. If $\varepsilon < z-y$ occurs with zero probability (measurement errors have a 'high enough' lower-bound), then the observer will never be misled into counting this individual as poor. This may be the case if the individual lies well above the poverty line, so that $z-y$ is too low, below the ε lower-bound.

More formally, for $z-y<0$,

$$\begin{aligned}y^*_{NP} = z \Longrightarrow \text{E}[\tilde{y}^*] - y^*_{NP} &= \int_{-\infty}^{z-y} \varepsilon\phi_{(\varepsilon)}\text{d}\varepsilon - (z-y)\Phi_{(z-y)}\\ &= \int_{-\infty}^{z-y} \varepsilon\phi_{(\varepsilon)}\text{d}\varepsilon - \int_{-\infty}^{z-y} (z-y)\phi_{(\varepsilon)}\text{d}\varepsilon\\ &= \int_{-\infty}^{z-y} \left[\varepsilon-(z-y)\right]\phi_{(\varepsilon)}\text{d}\varepsilon\end{aligned}$$

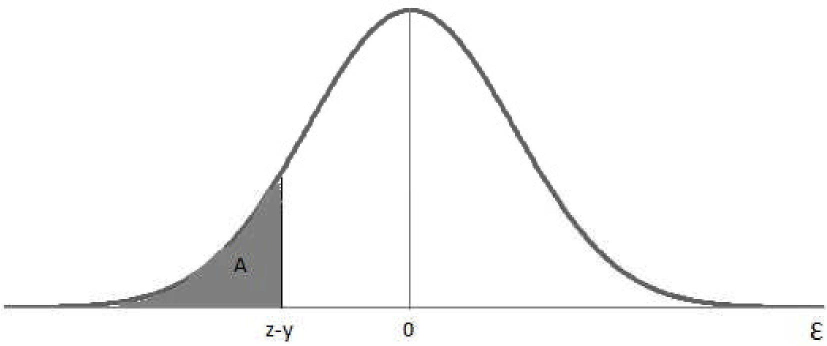
For later notational convenience, we define

$$\Delta \equiv \int_{-\infty}^{z-y} \left[\varepsilon-(z-y)\right]\phi_{(\varepsilon)}\text{d}\varepsilon$$

Since $[\varepsilon-(z-y)] < 0$ for all ε within the range of the integral, $\Delta \leq 0$. Hence,

$$\text{E}[\tilde{y}^*] - y^*_{NP} \leq 0$$

Figure 1: Effect of ε on the non-poor



For the inequality to turn into a strict inequality ($\mathrm{E}[\tilde{y}^*] < y^*_{NP}$), $\phi_{(\varepsilon)} > 0$ for at least some $\varepsilon < z-y$. If the probability of such 'strongly negative' measurement errors is positive, then poverty estimates will feed on underestimated censored outcomes, and hence poverty will be biased *upwards*.

To illustrate, Figure 1 assumes ε is drawn from a normal distribution. There, only error realisations below $z-y$ are consequential. The dark-shaded area A shows the probability of such type-II errors occurring.

2.2 The case of poor [P] individuals: $[z-y>0]$

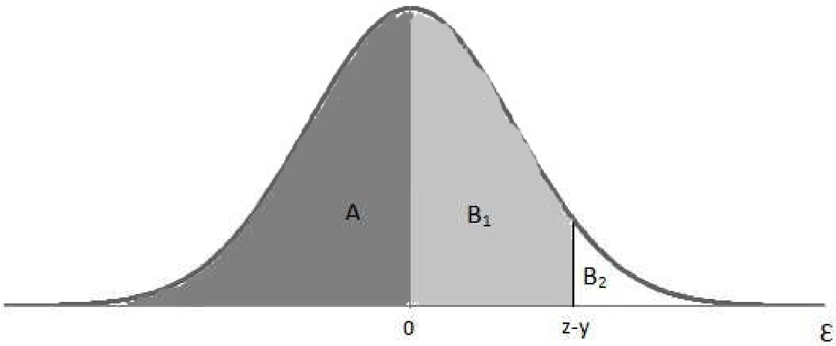
Take now an individual in actual poverty ($z-y>0$):

$$\begin{aligned} y^*_P = y \Longrightarrow \mathrm{E}[\tilde{y}^*] - y^*_P &= \int_{-\infty}^{z-y} \varepsilon\phi_{(\varepsilon)}\mathrm{d}\varepsilon + (z-y)\left[1-\Phi_{(z-y)}\right] \\ &= \int_{-\infty}^{z-y} \left[\varepsilon-(z-y)\right]\phi_{(\varepsilon)}\mathrm{d}\varepsilon + [z-y] \quad = \Delta + [z-y] \end{aligned}$$

Here, the possibility of measurement errors exacerbating observed poverty still exists, as captured again by Δ, as above. However, the last term $[z-y] > 0$ acts as a counterweight. In the case of truly poor individuals, measurement errors may induce the observer to take them as non-poor or underestimate the depth of their poverty (type I). The misalignment of $\tilde{y}^*$ with respect to y^*_P can thus go either way, upwards or downwards.

Figure 2 illustrates this, again under the assumption of normality. The dark-shaded area A (where $\varepsilon < 0$) is again related to cases where errors deceive the

Figure 2: Effect of ε on the poor



observer into overrating the outcome shortfall with respect to z. A new light-shaded area $B1$ (where $0 < \varepsilon < z - y$) shows errors reducing the perceived shortfall, but not 'positive enough' to turn the individual into an apparent non-poor. The observer is only led into this type-I fallacy if $\varepsilon > z - y$(area $B2$).

All three cases are implicit in $\mathrm{E}[\tilde{y}^*] - y_P^*$ above, which alternatively may be rewritten as follows, so as to unpack the elements corresponding to areas A, $B1$ and $B2$:

$$
\begin{aligned}
\mathrm{E}[\tilde{y}^*] - y_P^* &= \int_{-\infty}^{z-y} [\varepsilon - (z-y)]\, \phi_{(\varepsilon)} \mathrm{d}\varepsilon + [z-y] \\
&= \int_{-\infty}^{0} \varepsilon \phi_{(\varepsilon)} \mathrm{d}\varepsilon + \int_{0}^{z-y} \varepsilon \phi_{(\varepsilon)} \mathrm{d}\varepsilon + [z-y] \left[1 - \Phi_{(z-y)}\right]
\end{aligned}
$$

For later use, we use the following table to summarise. It shows how in both the P and the NP cases, measurement errors cause $\mathrm{E}[\tilde{y}^*]$ to deviate from y^*:

Two cases for $\mathrm{E}[\tilde{y}^*] - y_{i1}^*$

(P-case) $y < z$	$\Delta + [z-y]$
(NP-case) $y > z$	Δ

2.3 The results in Chesher and Schluter (2002)

In a given sample, some individuals are free of poverty (and their poverty can only be overestimated, as seen in case NP above), while others will not escape it (and their poverty may be over- or underrated, as in case P). The impact of measurement errors on aggregate poverty is hence ambiguous so far, with the distribution of true y outcomes across the sample as a determinant of any final bias. A second crucial determinant exists. Any deviation of $\tilde{y}^*$ from y^* will translate into a bias in poverty estimates according to how the specific poverty index depends on the censored outcome y^*, e.g. in the FGT family, $P_{(y^*;\alpha)} = \left(\frac{z-y^*}{z}\right)^\alpha$ (Foster et al. 1984).

Chesher and Schluter formalised the net effect for the particular case of FGT poverty measures. Using *multiplicative* measurement errors, they show that both the poverty gap ($\alpha = 1$) and the intensity of poverty ($\alpha = 2$) typically *increase* in the variance of measurement errors.[3] In those cases, the net effect is thus an *upward bias* in poverty estimates – negative measurement errors in y (which induce type-II errors, i.e. turn the non-poor into observationally poor and seemingly increase the gap and intensity of the truly poor) prevail over positive errors, which are ignored if the individual lies above the poverty line or are censored at $[z - y]$ as they 'reduce' the poverty of the truly poor. This asymmetry in the effects of positive and negative realisations of the measurement error underlies the final result. Of course, for $\alpha = 2$, the greater weight of downward deviations from the actual outcome y reinforces the effect.

In the case of the headcount ($\alpha = 0$), only strict type-I and type-II errors (entries into and exits from the observed poverty count) matter, and fictitious increases or decreases in gaps and intensities have no bearing. Thus, Chesher and Schluter find a weaker conclusion – the poverty headcount will increase in the variance of measurement errors if the density of outcome y is *rising* around the poverty line (e.g., if the poverty line lies below the mode value for y). Intuitively, this condition suggests that strict type-II errors (individuals just above z with $y + \varepsilon < z$) will in fact be more common than strict type-I errors, and observed poverty will thus increase.

[3]Since they resort to second-order Taylor approximations, some departures from this result may exist.

3 Multidimensional poverty and the dual cut-off

We now turn to multidimensional poverty, as conceptualised by the Alkire-Foster measure P^{AF}. In particular, a crucial defining feature of this measure is its dual cut-off strategy. Formally,

$$P_{(\alpha)}^{AF} = \frac{1}{nD} \sum_{i=1}^{n} \sum_{d=1}^{D} \max\left(0, 1 - \frac{y_{id}}{z_d}\right)^{\alpha} \mathbb{I}\left[\sum_{d=1}^{D} \mathbb{I}\left[y_{id} < z_d\right] > k\right]$$

where D and n are the numbers of the dimensions and individuals, respectively. Subscripts i and d stand hereafter for the i-th individual and the d-th dimension. The $P_{(\alpha)}^{AF}$ family is thus an adjusted version of the family of FGT unidimensional measures. The individual needs to suffer poverty in at least k dimensions for her unidimensional shortfalls (raised to power α) to add to the aggregate multidimensional index.

To make our point, we address here the particular case of $D = 2$. For the dual cut-off to be meaningful, we set $k = 2$. In terms of our discussion of biases in observed outcomes, this implies that the i-th individual counts as multidimensionally poor only if both $y_{i1} < z_1$ and $y_{i2} < z_2$. Hence,

$$y_{i1}^{**} = \mathbb{I}[y_{i2} < z_2]\text{Min}[y_{i1}, z_1] + (1 - \mathbb{I}[y_{i2} < z_2])\, z_1$$

where the double-star superscript denotes the censoring of an outcome entering a dual cut-off index, and $\mathbb{I}$ is an indicator function. It will be useful to think of y_{i1}^{**} in terms of four possible cases:

Four cases for y_{i1}^{**}

	$y_{i2} < z_2$	$y_{i2} > z_2$
$y_{i1} < z_1$	$y_{i1}^{**} = y_{i1}$	$y_{i1}^{**} = z_1$
$y_{i1} > z_1$	$y_{i1}^{**} = z_1$	$y_{i1}^{**} = z_1$

Unless the individual faces poverty in all $k = 2$ dimensions, unidimensional poverty is ignored, i.e. P^{AF} will treat the outcome in both dimensions as though the poverty line had been secured. To put it differently, if either dimension (say dimension 2) reaches the poverty line, then the other (say dimension 1) is inexorably censored (at z_1), regardless of y_{i1}.

Importantly, this definition of y_{i1}^{**} implies that type-II errors in one dimension will have no effect unless poverty is observed in the other dimension as well.

On the other hand, strict type-I errors (fictitious exits from poverty) will affect measured P^{AF} regardless of observed outcomes in the other dimension. This new asymmetry suggests that the competing forces described in section 2 are less likely to finally yield an upward bias in poverty estimates – recall type-I errors act in the opposite direction, and the second cut-off increases their relative weight.

To formalise, we define $\tilde{y}_{i1}^{**}$ which will be sensitive to measurement errors (ε_{i1} and, for the second dimension, ε_{i2}):

$$
\begin{aligned}
\tilde{y}_{i1}^{**} &= \mathbb{I}[\tilde{y}_{i2} < z_2]\text{Min}[\tilde{y}_{i1}, z_1] + (1 - \mathbb{I}[\tilde{y}_{i2} < z_2])\, z_1 \\
&= \mathbb{I}[\varepsilon_{i2} < z_2 - y_{i2}]\, (y_{i1} + \text{Min}[\varepsilon_{i1}, z_1 - y_{i1}]) + (1 - \mathbb{I}[\varepsilon_{i2} < z_2 - y_{i2}])\, z_1
\end{aligned}
$$

Using $\Psi_{(\varepsilon_{i2})}$ to denote the c.d.f. for the measurement error for ε_{i2}, the expected value can be written as follows:

$$
\begin{aligned}
\text{E}\left[\tilde{y}_{i1}^{**}\right] = \Psi_{(z_2 - y_{i2})} &\left(y_{i1} + \int_{-\infty}^{z_1 - y_{i1}} \varepsilon \phi_{(\varepsilon)} \text{d}\varepsilon + [z_1 - y_{i1}] \left[1 - \Phi_{(z_i - y_{i1})}\right] \right) \\
&+ \left[1 - \Psi_{(z_2 - y_{i2})}\right] z_1
\end{aligned}
$$

We can now spell out the expected deviation between observed, censored outcomes and their true values ($\text{E}\left[\tilde{y}_{i1}^{**}\right] - y_{i1}^{**}$). Again, it will be useful to be explicit about the four possible cases (defined by *true* poverty conditions in either dimension):

Four cases for $\text{E}\left[\tilde{y}_{i1}^{**}\right] - y_{i1}^{**}$

	$y_{i2} < z_2$	$y_{i2} > z_2$
$y_{i1} < z_1$	$\Psi_{(z_2 - y_{i2})}\Delta + (z_1 - y_{i1})$	$\Psi_{(z_2 - y_{i2})}\Delta$
$y_{i1} > z_1$	$\Psi_{(z_2 - y_{i2})}\Delta$	$\Psi_{(z_2 - y_{i2})}\Delta$

where again $\Delta = \int_{-\infty}^{z_1 - y_{i1}} [\varepsilon - (z_1 - y_{i1})]\, \phi_{(\varepsilon)} \text{d}\varepsilon$. Recall $\Delta = \text{E}\left[\tilde{y}_{i1}^{*}\right] - y_{i1}^{*}$ in the $[y_{i1} > z_1]$ unidimensional case, and also $\Delta \leq 0$. In the dual cut-off, multidimensional setting, the negative impact of Δ is weakened in all four cases by $\Psi_{(z_2 - y_{i2})}$, the probability of observing poverty in the other dimension. As expected, the driver of the negative (positive) bias in outcomes (poverty estimates) [see section 2] is now conditioned to the other dimension.

A counteracting effect exists. In the ($y_{i1} < z_1$, $y_{i2} > z_2$) case, the term $[z_1 - y_{i1}]$ disappears, if compared with the unidimensional case. Type-I errors now have no bearing if the other dimension is free of poverty. Whether the dual cut-off finally mitigates or exacerbates the impact of measurement errors remains an empirical issue. Moreover, since true y_{id} values are unobserved and thus the actual distribution of individuals into the four cases above ($y_{id} \gtrless z_d$) will remain unknown, the attenuation or exacerbation of the bias can only count as presumptions. However, since the attenuation effect of $\Psi_{(z_2-y_{i2})}$ is at work in all four cases above, only a strong relative concentration of individuals in the ($y_{i1} < z_1$, $y_{i2} > z_2$) case would reverse *our presumption that the dual cut-off makes P^{AF} more robust to measurement errors.* In empirical work, this presumption will be further strengthened if outcomes are (as usual) positively correlated across dimensions, since this would suggest individuals do not agglomerate in off-diagonal cases (such as $y_{i1} < z_1$, $y_{i2} > z_2$).[4] All in all, and with the necessary caveats in mind, we do expect a weaker upward-bias in P^{AF}, as compared with unidimensional indices.

4 Empirical exercise

A direct test of the biases discussed above would require a dataset providing both true and ε-distorted outcomes. Needless to say, this would be unusual. For this reason, in this section we follow an indirect route to capture the traces of measurement errors in a standard LSMS dataset. In particular, we use 2010 LSMS data from Peru, which provide no special information on the accuracy of reported wellbeing outcomes but, as any other typical dataset, it does identify the informant for each part of the questionnaire. To exploit this piece of information, we note that besides measurement errors, no other reason exists for the identity of the informant to have an impact on reported outcomes.

4.1 The data

The dataset includes 83,373 individuals in 27,176 households surveyed from the whole country by the National Statistics Institute (INEI). Our choice to use INEI data from Peru is warranted by positive appraisals of past INEI surveys (e.g. World Bank, 2005), so that our discussion below does not seem to be driven by the presence of atypical, badly pervasive measurement errors. We take per capita consumption and schooling as our wellbeing outcomes.

[4]Note correlations of observed, ε-distorted outcomes are unbiased estimates of true correlations.

Both will be defined at the individual level, although per capita consumption will be common to all household members.

Consumption expenditures of the household are reported in great detail. The total consumption figure includes eight categories (each referring to a specific period prior to the interview), which in turn add together amounts spent on a plethora of specific items. For each non-individual category (say, food or water supply), a household member acted as informant. Table 1 shows that the spouse of the household head is the source of consumption information in 68% of cases.

Table 1 also shows that years of schooling are usually reported by the individual herself. When she does not self-report, as many as 51% of cases are reported by the household spouse. This rate increases to 70% when the sample is restricted to only households with both a household head and a spouse, and further only to those two individuals in the household (head and spouse) provided they are aged between 30 and 65. We confine the sample to this narrow group for two reasons.[5] First, schooling attainments (one of our wellbeing dimensions) of these individuals can be thought of as given.

Table 1: Response rates in consumption and schooling

Panel A: Consumption	Whole Sample	Our Sample
HH Head	27.00%	16.00%
Spouse	68.00%	84.00%
Son/daughter	4.09%	-
Others	0.91%	-
Panel B: Schooling	**Whole Sample**	**Our Sample**
Individual self-reported	70%	93%
Individual did not self-report	30%	7%
If she did not self-report, who did?		
HH Head	32.00%	29.00%
Spouse	51.00%	70.00%
Son/daughter	13.45%	-
Others	3.55%	-

Second, and more importantly, we wish to control for characteristics of *both* the head and his spouse as consumption predictors. Our strategy rests on

[5] Since we will not identify the determinants of consumption and schooling in order to draw policy implications, the external validity of our restricted sample is not an issue.

the intuition that the identity of the informant (typically, the head or the spouse) should have no true impact on consumption if the characteristics of both the head and his spouse are controlled as such. If the identity of the informant retains any explanatory power after these controls, we will read this as symptom of (biasing) measurement errors in the data, i.e. the head and his spouse are not equally reliable as information sources.

For consumption, the poverty line is calculated by INEI and included in the dataset.[6] We take complete primary education (6 years of schooling) as the education poverty line for our sample of individuals aged 30 or above.

4.2 Empirical strategy

We first estimate per capita consumption (y_{i1}) as a function of household characteristics (X_i^H). In the case of schooling (y_{i2}), we use individual characteristics (X_i^I) and, in particular, we use variables describing individual conditions in her youth, during her student years. In both cases, we add controls for characteristics (W_i) which may explain outcomes and also correlate with the choice of who will act as informant – in the absence of these controls, the informant variable would arguably pick other effects besides measurement accuracy.

Formally,

$$y_{i1} = \beta_0 + \beta_1 X_i^H + \beta_2 W_i + \varepsilon_{i1} + u_{i1}$$
$$y_{i2} = \gamma_0 + \gamma_1 X_i^I + \beta_2 W_i + \varepsilon_{i2} + u_{i2}$$

where measurement errors ε_{id} can be decomposed as

$$\varepsilon_{i1} = R_{i1}\varepsilon_{i1}^{\text{head}} + (1 - R_{i1})\,\varepsilon_{i1}^{\text{spouse}}$$
$$\varepsilon_{i2} = R_{i2}\varepsilon_{i2}^{\text{self}} + (1 - R_{i2})\,\varepsilon_{i2}^{\text{other}}$$

where R_{i1} indicates whether the household head reported expenditures and R_{i2} indicates whether the individual reported her own years of schooling. This decomposition captures the notion that each information source is fraught with a specific form of measurement error. While we assume errors are zero-mean in all cases ($\text{E}\left[\varepsilon_{i1}^{\text{head}}\right] = \text{E}\left[\varepsilon_{i1}^{\text{spouse}}\right] = \text{E}\left[\varepsilon_{i2}^{\text{self}}\right] = \text{E}\left[\varepsilon_{i2}^{\text{other}}\right] = 0$), each case

[6]This is an absolute poverty line, based as usual on minimal nutritional intakes and Engel coefficients.

may be characterised with a different variance. Some informants are more accurate than others. Presumably,

$$\text{Var}\left[\varepsilon_{i1}^{\text{head}}\right] > \text{Var}\left[\varepsilon_{i1}^{\text{spouse}}\right]$$

and

$$\text{Var}\left[\varepsilon_{i2}^{\text{self}}\right] < \text{Var}\left[\varepsilon_{i2}^{\text{other}}\right]$$

Empirically, we simply rewrite the equations for y_{i1} and y_{i2} as follows:

$$y_{i1} = \beta_0 + \beta_1 X_i^H + \beta_2 W_i + \tau_1 R_{i1} + \varepsilon_{i1}^{\text{spouse}} + u_{i1}, \text{ where } \tau_1 = \varepsilon_{i1}^{\text{head}} - \varepsilon_{i1}^{\text{spouse}}$$
$$y_{i2} = \gamma_0 + \gamma_1 X_i^I + \beta_2 W_i + \tau_2 R_{i2} + \varepsilon_{i2}^{\text{other}} + u_{i2}, \text{ where } \tau_2 = \varepsilon_{i1}^{\text{self}} - \varepsilon_{i1}^{\text{other}}$$

While both τ_1 and τ_2 are thus random coefficients, for the sake of simplicity we estimate both models by standard OLS, except we instrument R_{i2}. Whether the individual is able to answer the questionnaire herself is not entirely exogenous, since, e.g. working conditions may imply frequent absences from the family home. For this reason, we take as instruments whether the family was interviewed on a Sunday or in the summer, and both instruments are interacted with the sex of the individual. In the case of consumption (R_{i1}), endogeneity fears are less compelling, since the outcome variable is defined at the household level.

To sum up, we will estimate the impact of the informant choice on wellbeing outcomes and interpret it as the consequence of differences in the magnitude of measurement errors. *A priori*, since all errors are zero-mean, we expect no significant impact on consumption and schooling levels (provided W_i variables are properly controlled for).

However, following our discussion in section 2, we do expect that poverty indices should react to a change in the information source – measurement errors drawn from a distribution with greater variance should lead to a greater (upward) bias in poverty estimates. This should be reflected in significant coefficients for R_{i1} (and R_{i2}) if y_{i1} (and y_{i2}) are replaced by unidimensional poverty indices based on those levels. Finally, since the dual cut-off in P_i^{AF} should act as an attenuation device of the upward bias in poverty estimates (see section 3), we expect the coefficients for R_{i1} and R_{i2} to lose significance as multidimensional poverty is placed on the LHS.

4.3 Results

Table 2 reports our results for consumption and schooling levels. In neither case is the effect of the identity of the informant statistically different from zero, even though the negative impact of self-reporting on schooling is close to significant. The latter result raises a caveat regarding our results below, since it suggests the information source variable may be capturing some effects beyond (zero-mean) measurement errors, albeit weakly. In our schooling results, the inclusion of both family size and marital status (both significant) may only partially explain away individual characteristics related both to schooling and the probability of answering the survey interview (W).

All controls exhibit the expected signs. Age and schooling (for the head and spouse) act as predictors of per capita consumption (X_i^H). Individual characteristics predicting schooling (X_i^I) also include age, as well as sex and their interaction.[7] To capture conditions early in the life of the individual, we add whether Spanish was her native language and the rate of individuals of the same cohort (30 to 65) completing at least primary school.

As levels are replaced by FGT poverty indices on the LHS, controls retain significance and change their signs, as expected. They are not reported on Table 3, which focuses on coefficients for R_{i1} and R_{i2}. Our results are in keeping with our discussion above on the impact of measurement errors. Turning to unidimensional poverty first, if the household head (rather than his spouse) reports expenditures, his lower accuracy results in an upward shift in observed FGT1 and FGT2. Likewise, if someone other than the individual reports her schooling, the (presumably) greater measurement error raises observed poverty. In the case of the FGT0, both consumption and schooling poverty headcounts are unaffected by the information source. This is consistent with the results in Chesher and Schluter, where the impact on FGT0 is conditional on the actual distribution of true outcomes y_{i1} and y_{i2} around their poverty lines.[8]

Importantly, Table 3 shows that P_i^{AF} is in no case affected by changes in the identity of the informants. As expected from the discussion in section 3, this dual cut-off measure seems to withstand more robustly the biasing effect of measurement errors. As a robustness check, we omitted each R_{i1} and R_{i2} in turn and found no change in the general pattern – the coefficient on the remaining information source variable is in no case statistically different from zero.

[7] X_i^H excludes sex of the household head because all households are male-headed.

[8] Through a different route, McGarry (1995) estimates that her poverty headcounts are also upward biased due to measurement errors.

Table 2: Wellbeing outcome levels

	Log(Consumption) OLS	Schooling IV
R_1: HH Head reported expenditures	-0.0049	—
	(0.0076)	
R_2: Someone else reported schooling	—	-3.633
		(2.621)
W: Log(Family size)	-0.4078***	-1.1656***
	(0.0093)	(0.7879)
W: Marriage	0.0843***	0.9261***
	(0.0071)	(0.0619)
X^H: HH Head Age	0.0067***	—
	(0.0006)	
X^H: HH Head Schooling	0.0319***	—
	(0.0010)	
X^H: Spouse Age	0.0068***	—
	(0.0007)	
X^H: Spouse Schooling	0.0368***	—
	(0.0009)	
X^I: Male	—	-0.1003
		(0.4929)
X^I: Age	—	-0.1325***
		(0.0060)
X^I: Male$\times$Age	—	0.3607***
		(0.0085)
X^I: Native Spanish language	—	3.4216***
		(0.0851)
X^I: Primary school (sex-region rates)	—	7.6466***
		(0.8262)
Region dummies	yes	yes
Observations	18,856	18,857
R-squared	0.57	0.36

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust errors are reported in parentheses.

Table 3: Information source and poverty indices

	Consumption	Schooling IV	Multidim P^{AF} IV
FGT0 [Probit]			
HH Head reported expenditures	-0.0229	—	1.1188
	(0.2673)		(1.3708)
Someone else reported schooling	—	1.1257	-0.0004
		(0.9228)	(0.0657)
FGT1 [Tobit]			
HH Head reported expenditures	0.1267*	—	0.5848
	(0.0069)		(0.6015)
Someone else reported schooling	—	1.2195*	-0.0075
		(0.7227)	(0.0256)
FGT2 [Tobit]			
HH Head reported expenditures	0.0846**	—	0.5114
	(0.0036)		(0.4217)
Someone else reported schooling	—	1.2430*	-0.0084
		(0.6339)	(0.0182)

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust errors are reported in parentheses. Log-consumption and schooling regressions include RHS variables as on Table 2. The regressions for P_i^{AF} includes both sets of regressors X_i^H and X_i^I.

5 Concluding remarks

While an awareness of measurement errors should be inherent in any inference from empirical poverty figures, this caveat passes largely unnoticed. In this paper, we first discuss formally the impact of measurement errors on unidimensional poverty estimates, and we then extend this analysis onto the multidimensional framework. In particular, we find that the likely upward bias of unidimensional measures is attenuated when the multidimensional index is based on a dual cut-off strategy, such as P^{AF}. We cannot generalise this result, chiefly because a high proportion of 'off-diagonal' individuals (poor in one dimension and non-poor in the other) might overturn it if combined with strong type-I errors. Nonetheless, the attenuation effect remains in our view the likely product of the dual cut-off. A brief empirical exercise with data from Peru suggests this as well.

Of course, the shift from a uni- to a multidimensional standpoint raises other issues with regards to measurement errors. For instance, some outcomes might be more vulnerable to measurement errors than consumption, so that adding new dimensions to the conventional consumption-based analysis might in practice reduce the accuracy of estimates. The dual cut-off would then act as a counterweight.

Finally, further formalising the role of a dual cut-off strategy might shed light on the advantages of applying it to other settings, e.g. to the assessment of poverty over time, as in Foster (2009). This remains an avenue for further work.

References

ALKIRE, S. (2002). “Dimensions of Human Development”, *World Development*, 30(2), 181-205.

ALKIRE, S. and J. FOSTER (2011). “Counting and Multidimensional Poverty Measurement”, *Journal of Public Economics*, 95(7-8), 476-487.

ATKINSON, T. and F. BOURGUIGNON. “The Comparison of Multi-Dimensional Distribution of Economic Status”, *The Review of Economic Studies*, 49(2), 183-201.

BOURGUIGNON, F. and S. Chakravarty (2003). “The Measurement of Multidimensional Poverty”, *Journal of Economic Inequality*, 1(1), 25-49.

CHESHER, A. and C. SCHLUTER (2002). “Welfare Measurement and Measurement Error”, *The Review of Economic Studies*, 69(2), 357-378.

DEUTSCH, J. and J. Silber (2005). “Measuring Multidimensional Poverty. An Empirical Comparison of Various Approaches”, *The Review of Income and Wealth*, 51(1), 145-174.

FOSTER, J. (2009). “A Class of Chronic Poverty Measures”, in T. Addison, D. Hulme and R. Kanbur (eds), *Poverty Dynamics: Interdisciplinary Perspectives*, Oxford University Press.

MCGARRY, K. (1995), “Measurement Error and Poverty Rates of Widows”, *The Journal of Human Resources*, 30(1), 113-134.

RAVALLION, M. (2011). “On Multidimensional Indices of Poverty”, *Journal of Income Inequality*, 9(2), 235-248.

WORLD BANK (2005). *Peru. Opportunities for All*, Report 29825-PE.