Economic Mobility and Expenditure Growth:

The Effects of Measurement Error in Peruvian Data

Renato Ravina

April, 2008

I. Introduction

Although several authors have shown evidence that expenditure and income data obtained from household surveys is typically measured with a large amount of error[1], there are not many studies using data from Latin America that address this issue. As a result, studies that calculate different economic indicators, such as poverty or inequality indexes or economic mobility for Latin America countries, are usually contaminated by measurement error in the data (Duval Hernández et.al., 2006).

This paper uses nationally representative panel data from Peru to estimate the effect of measurement error on estimated economic mobility and the growth rate of per capita expenditures for years 2004 to 2006. More specifically, two alternative methods proposed by Luttmer (2002) and Glewwe and Dang (2005) are implemented to estimate the impact of the measurement error on the observed variance of the logarithm of per capita expenditures. These results are then used to assess the relative importance of measurement error in the estimation of economic mobility in Peru between 2004 and 2006 by simulating the join distribution of the logarithm of per capita expenditures corrected for the effect of measurement error. Finally, the magnitude of the bias in the growth of rate of per capita expenditures by quintiles is also evaluated.

One of the main findings of this paper is that conservative estimates suggest that measurement error in per capita expenditures accounts for between 11 and 13 percent of the observed

---

[1] See for instance Bound and Krueger (1991) and Pischke (1995)

variance of per capita expenditures. Also, economic mobility indices estimated from observed data are upward biased by 24 to 50 percent. Finally, the results suggest that measurement error in per capita expenditure produces an overestimation of its growth rate that ranges from 46 to 61 percent for the households in the first quintile, and from 39 to 43 percent for the second quintile. For the third quintile the overestimation in the growth rate of per capita expenditures is of about 40 percent. For the fourth quintile, the overestimation varies between 14 and 19 percent. Finally, results for the fifth quintile suggest that the direction of the bias depends on the assumptions about the variances of the logarithm of per capita expenditures.

The remaining sections of this paper are organized as follows: Section II presents a brief review of the more recent literature involving the assessment of the effects of measurement error on economic mobility and mean of expenditures. In Section III, the methods applied in this paper to estimate such effects are explained. Section IV describes the data to be used. Section V provides and discusses the results. Finally, the last section presents the main conclusions of the study.

II. Literature Review

There are several possible sources of error in expenditure data from household surveys.[2] According the typology of non-sampling error in expenditure data provided by Neter (1970), the following groups can be distinguished: (i) recall errors associated with the fading of people's memories; (ii) telescoping errors caused by the incorrect placing of certain expenditure

[2] See Deaton (2000) for an extensive discussion.

in time by respondents; (iii) reporting errors associated with respondents being overwhelmed either by the length of the survey or by the number of items covered; (iv) under- or over-reporting associated with the "prestige" of the item; (v) conditioning effects associated with repeated interviewing; (vi) respondent effects associated with the particular member of the household who answers the questionnaire; (vii) interviewer effects; and (viii) effects associated with the design of the instrument. To these categories, Deaton (2000) adds also biases in the data due to non-responses or to the use of an inadequate sampling frame.

Several methods have been proposed to remove the effects of measurement error on income or expenditure data. One of them is based on the so-called validation studies. The idea is to compare household survey data to other sources of information about income or expenditures, which usually come from administrative data.[3] For instance, Bound and Krueger (1991) compare reported income in the U.S. from Current Population Survey data to employer-reported Social Security earning records. They find that the measurement error of reported earnings is positively auto correlated and negatively correlated with true earnings. Pischke (1995) obtains similar results using U.S. data from the Panel Study of Income Dynamics Validation. Finally, Battistin (2002) uses data from household survey to data from diary surveys (that are completed by the respondents) from the U.S. Consumer Expenditure Survey to measure the extent of measurement error in expenditure data. He obtains some evidence of measurement error affecting the aggregate measure of consumption both for diary and recall-based data. In general, it is not possible to use validation studies based on survey data

[3] See Bound, Brown, and Mathiowetz (2001) for an extensive review.

measured with error in developing countries due to the unavailability of administrative data that can be matched to household surveys.

Regarding the effects of measurement error on the estimation of economic mobility, Antman and Mckenzie (2005) propose to construct pseudo panels from cross-sectional surveys. Instead of tracking individual observations, pseudo panels allow analysis of economic mobility for cohorts of individuals, which can be based on age, gender, education, etc. The method then uses cohort means within each time period in order to eliminate individual-level measurement error.

The pseudo-panel approach had its critics, such as see Duval Hernández ***et.al***. (2006) and Deaton (1997). First, this method might still lead to biases if there is time-varying cohort-level measurement error. Also the pseudo-panel analysis can entail certain biases when it fails to track a consistent group of individuals over time, due to events like migration, deaths, and household dissolution and creation. Finally, switching the analysis from individual or household expenditure to average cohort expenditure eliminates the possibility of studying any intra-cohort income mobility.

An alternative to overcome the measurement error problem is based on the instrumental variable methodology. In particular, Glewwe and Dang (2005) propose to use second measures of expenditures or variables that are caused by expenditures as instruments. The method requires one instrument and panel data that captures households at two different points of time. Another instrumental variable approach has been proposed by Luttmer (2002). His method allows researchers to estimate the impact of measurement error in expenditures without panel data, although it requires at least two instruments. In this paper both Glewwe and Dang's and

Luttmer's methods are used and their results are compared. The details of the methodological procedures are described in the following section.

III. Methodology

Measurement error

In order to estimate the magnitude of the bias caused by measurement error in the variance of Peruvian households' expenditures, two distinct methods are used. First, following the method developed by Luttmer (2002), if two instruments for per capita expenditures are available, it is possible to estimate the impact of measurement error for one particular year using cross sectional data. Assume that the measured per capita expenditures are decomposed into a corrected component and an orthogonal measurement error:

$$E = E^* + m \tag{1}$$

where $\boldsymbol{E}$ is the observed per capita expenditures in the households, $\boldsymbol{E^*}$ are the true unobserved per capita expenditures. Also assume that the measurement error m has a variance of s_m^2.

Now let $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$ be two instruments for $\boldsymbol{E^*}$. For these instruments to be valid, the following assumptions are needed:

$$E[z_i m] = 0,\ i = 1,2 \tag{2}$$

$$r(z_i, E*) \neq 0, i = 1,2 \tag{3}$$

Equation 2 implies that both instruments are uncorrelated with the measurement error in Equation 1. On the other hand, Equation 3 implies that both instruments are correlated with ***E*,*** and thus the correlation coefficient of each instrument and ***E***** is different than zero.[4] Further, assume that the variances of the error terms of the regressions of ***E**** on z_1 and ***E**** on z_2 are s_v and s_w, respectively. We can then use the instrumental equations to estimate the variance of the real per capita expenditures, s^2_{E*}.

Solving for the covariances between measured per capita expenditure and the two instruments, and between both instruments, Luttmer (2002) obtains:

$$Cov[x, z_1] \equiv s_{xz_1} = a_1 s^2_{x*} \tag{4}$$

$$Cov[x, z_2] \equiv s_{xz_2} = g_1 s^2_{x*} \tag{5}$$

$$Cov[z_1, z_2] \equiv s_{z_1 z_2} = a_1 g_1 s^2_{x*} \tag{6}$$

where a_1 is the slope parameter of the regression of z_1 on $\boldsymbol{x}$ and g_1 is the slope parameter of the regression of z_2 on $\boldsymbol{x}$. Equations (4), (5) and (6) are then solved for s^2_{x*}:

[4] In a more general setting, the instrumental variables method requires the instruments to be partially correlated with the endogenous variable. Since Equation 1 only includes one explanatory variable, it is possible to reduce more complicated rank conditions to the assumption made in Equation 3. See Wooldridge (2002).

$$\sigma_{x^*}^{2} = \frac{\sigma_{xz_1}\,\sigma_{xz_2}}{\sigma_{z_1 z_2}} \tag{7}$$

Having found the variance of the logarithm of real per capita expenditures, the variance of the logarithm of the measurement error can be obtained as follows:

$$\sigma_{e}^{2} = \sigma_{x}^{2} - \sigma_{x^*}^{2} \tag{8}$$

The first instrument used by Luttmer (2002) is based on a question in which responders rank their standard of living using a discrete scale. We use a somewhat similar instrument, in the sense that it is also based on households' own subjective perception of their economic situation. More specifically, we use the subjective minimum income per capita, which is the ratio between the minimum amount of money required by the household to make ends meet each month, according to the head of the household's subjective assessment, and the number of members of the household. This variable, although highly correlated with household per capita expenditure, should not be correlated with the measurement error in measured per capita expenditure.

The second instrument, which is also used by Luttmer (2002), is income per capita. This instrument is also highly correlated with household expenditure. However, we can expect that the measurement error of income per capita is correlated with the measurement error of per capita expenditure (i.e. households that underreport income are likely to underreport expenditures as well). We then use the logarithm of household income per capita as one of our instruments due to the lack of a better alternative.

As argued by Luttmer (2002), even though using the logarithm of income as an instrument would lead us to obtain biased estimators of s^{2}_{x*} and s^{2}_{e}, it is possible to determine the direction of those biases. Thus, the expected positive correlation between the measurement error of income and the measurement error of expenditure would cause an upward bias in s_{xz_2}. This would lead us to overestimate the fraction of the variance of observed household income that can be explained by its true variance and underestimate the variance of the measurement error. Hence, we should interpret all our estimates of s^{2}_{e} as lower bounds of the variance of the measurement error in the logarithm of household expenditures.

The second approach for estimating the importance of the measurement error in household surveys' expenditure data is based on the method presented in Glewwe and Dang (2005), which requires panel data that includes the same households at two different points of time. This method starts by assuming that the measured household per capita expenditure in period 1, ***x***, is the product of the true unobserved household per capita expenditure, ***x****, and a random measurement error, $e_x > 0$, which in turn implies:

$$\log(x) = \log(x*) + \log(e_x) \tag{9}$$

Assuming that both the unobserved per capita expenditure and the error term at period 1 follow lognormal distributions implies that the observed per capita expenditure must also follow a lognormal distribution, such that:

$$\log(x) \sim N\left(l_x, s^{2}_{x*} + s^{2}_{ex}\right)$$

where l_x is the mean of ***x***, s^2_{x*} is the variance of ***x**** and s^2_{ex} is the variance of e. Further, assume that all these assumptions also hold for per capita expenditures at period 2, ***y***, such that:

$$\log(y) = \log(y^*) + \log(e_y) \tag{10}$$

$$\log(y) \sim N\left(l_y, s^2_{y*} + s^2_{ey}\right)$$

where ***y*** is the measured per capita expenditure in period 2, ***y**** is the real (unobserved) per capita expenditure, $e_y > 0$ is the random measurement error in per capita expenditures at period 2, l_y is the mean of ***y***, s^2_{y*} is the variance of ***x**** and s^2_{ey} is the variance of e_y.

In order to estimate the importance of measurement error in per capita expenditures Glewwe and Dang (2005) rely on the following equations:

$$\log(x^*) = a_1^* + b_1^* \log(y^*) + u_1 \tag{11}$$

$$\log(y^*) = a_2^* + b_2^* \log(x^*) + u_2 \tag{12}$$

Since both ***x**** and ***y**** are unobservable, instrumental variables must be used to obtain consistent estimations of the parameters in Equations 11 and 12. Based on derivations provided by Glewwe (2005), Glewwe and Dang argue that if the instrument used is a variable caused by ***x**** ***(y*)***, then one can obtain unbiased estimates of b_1^* (b_2^*).

In this study, the subjective minimum income per capita is used as an instrument of the logarithm of per capita expenditures. As previously explained, although this variable is highly correlated with household per capita expenditure, it should not be correlated with the measurement error in observed per capita expenditure. Moreover, this variable is arguably caused by per capita expenditures, since households that spend more tend to require greater levels of income to pay their expenditures. In contrast, we should not expect that the subjective minimum income per capita causes per capita expenditure, because expenditures should depend on household intertemporal preferences for goods and services and income level, but not on subjective minimum income.

Although it is not expected that current expenditure causes the current subjective minimum income, it can be argued that past levels of expenditure may cause the current subjective minimum income. If that is the case, the subjective minimum income would not be a good instrument to estimate Equation 11, since it would be correlated with $\boldsymbol{u_1}$**.** However, by assuming that the proportional contribution of the measurement error to the variance of ln(x) is equal to the proportional contribution of the measurement error to the variance of ln(y), it can be shown that:

$$\frac{b_1^*}{b_1} = \frac{Var[\log(y)]}{Var[\log(y^*)]} = \frac{Var[\log(x)]}{Var[\log(x^*)]} = \frac{b_2^*}{b_2} \tag{13}$$

where b_1 and b_2 are the Ordinary Least Squares estimates of the regression of ln(x) on ln(y) and of ln(y) on ln(x), respectively. The proportional variance assumption then implies:

$$b_1^* = b_1 \frac{b_2^*}{b_2} \tag{14}$$

Therefore, by obtaining consistent estimates of b_2^* using the minimum income per capita as an instrument in Equation 12, b_1 and b_2 by Ordinary Least Squares, ***Var[log(x)]*** and ***Var[log(y)]*** from observed data, we can estimate the variances of the logarithm of x* and y* with Equation 13 and b_1^* by Equation 14. Then we can finally estimate the variance of the logarithm of e_x and e_y using the following equations shown by Glewwe and Dang (2005):

$$Var[\log(e_x)] = Var[\log(x)] - Var[\log(x^*)] = Var[\log(x)]\left(1 - \frac{b_2^*}{b_2}\right) \quad (15)$$

$$Var[\log(e_y)] = Var[\log(y)] - Var[\log(y^*)] = Var[\log(y)]\left(1 - \frac{b_1^*}{b_1}\right) \quad (16)$$

Hence, we can estimate the variance of the logarithm of measurement error in per capita expenditures using either Luttmer (2002)'s method or Glewwe and Dang's (2005) method. For the former, cross sectional data cross sectional data may be used if two instruments for the (logarithm) of per capita expenditures are available. In contrast, Glewwe and Dang (2005)'s method requires only one instrument, which should be caused by per capita expenditures when a panel of households of at least two years is available. In this paper both approaches are implemented using Peruvian data in order to compare the estimates under two different sets of assumptions. These estimates then can be used to recover the distribution of the real per capita expenditures. This will be done in order to estimate economic mobility and the growth rate of per capita expenditures in Peru between 2004 and 2006.

Economic mobility and transition matrices

One of the consequences of household expenditures being measured with error is that it leads to overestimates of economic mobility over time. In order to evaluate the extent to which economic mobility is overestimated, we use the following index proposed by Glewwe (2005):

$$m(x_0, x_1) = 1 - r(\ln(x_0), \ln(x_1)) \tag{17}$$

where x_i the household expenditure in period ***i***, and $r(a,b)$ is the Pearson coefficient of correlation between ***a*** and ***b***, which is assumed to be positive. The idea here is that the larger the correlation of household expenditures over time, the lower the economic mobility. Since the coefficient of correlation is between 0 and 1, economic mobility will also be between 0 (no mobility) and 1 (complete mobility). As Glewwe (2005) points out, this particular index measures ***relative*** mobility, since it focuses on changes over time in the relative position of households in expenditures distribution. He also shows this index of economic mobility satisfies some desirable mathematical conditions.[5]

Another way to analyze economic mobility is to construct transition matrices for household expenditures between two periods of time. Although it does not allow us to get one unique measure of economic mobility, this constitutes a complement for the mobility index, since transition matrices are useful to evaluate how the expenditure distribution of households

[5] In particular, Glewwe (2005) shows that this mobility index is ***strongly relative*** (i.e. it focuses on changes in expenditure shares, not changes in expenditure over time) and satisfies the Atkinson-Bourguignon condition (i.e. measured economic mobility increases if a household with higher expenditures than a second household in both time periods switches its level of expenditures with the second household in one of the two periods).

changes among quintiles from one period to another. Thus, it allows us to study the percentage of households that go from one quintile to another compared to those that stay in the same quintile.

In this study both the economic mobility indices and the transition matrixes are estimated using two methods. First, the variance of the logarithm of the measurement error of per capita expenditures for years 2004 and 2006 is calculated following Luttmer's method, as previously described. Then, a similar method is applied to calculate the distribution of the changes in the logarithm of per capita expenditures. In particular, the variance of the change in per capita expenditures, $s^2_{\Delta x^*}$,can be found by:

$$s^2_{\Delta x^*} = \frac{s_{\Delta x \Delta z_1} s_{\Delta x \Delta z_2}}{s_{\Delta z_1 \Delta z_2}} \tag{18}$$

where the covariances $s_{\Delta x \Delta z_1}$, $s_{\Delta x \Delta z_2}$, and $s_{\Delta z_1 \Delta z_2}$ are defined analogously as in Equations 4, 5 and 6, using the variations for each variable. Thus the variance of the measurement of the change in per capita expenditures, $s^2_{\Delta e}$, is just the difference between the variance of the observed change in per capita expenditures, $s^2_{\Delta x}$, and $s^2_{\Delta x^*}$:

$$s^2_{\Delta e} = s^2_{\Delta x} - s^2_{\Delta x^*} \tag{19}$$

Using these estimates and those obtained by Equations 7 and 8, we can generate simulations of per capita expenditures in 2004 the variation per capita expenditures from 2004 to 2006 corrected by measurement error to calculate simulated per capita expenditures in 2006. Then, these simulations can be used to estimate economic mobility and generate transition matrixes. In this study, 50,000 simulations were generated following this method.

A second method to evaluate economic mobility implemented in this study is based on the estimation of the joint distribution of per capita expenditures in 2004 and 2006 using either Equation 11 or Equation 12 from Glewwe and Dang (2005). In order to obtain the joint distribution of per capita expenditures in 2004 and 2006, the linearity assumption in Equation 11 and Equation 12 needs to be checked. In our case, the linearity assumption holds for both equations (i.e. quadratic terms for the explanatory variables were found to be statistically insignificant). In order to follow a similar procedure to the one followed by Glewwe and Dang (2005), the joint distribution of per capita expenditures is obtained using Equation 12.

To begin, estimates of b_1^*, ***Var[ln(y*)]***, a_1^* and ***Var(u₁)*** are required**.** The parameter b_1^* can be estimated by Equation 14, once b_2^* has been previously estimated by Equation 12 using the instrumental variable approach previously explained. Then ***Var[ln(y*)]*** can be estimated using Equation 13. However, Glewwe and Dang (2005) shows that the parameter obtained should be interpreted as the upper bound of the variance of the logarithm of ***y****. Thus, at least two assumptions should be made about its value. Finally, the parameters a_1^* and ***Var(u₁)*** can be estimated using the following equations derived by Glewwe and Dang (2005):

$$a_1^* = \overline{\ln(x)} - b_1^* \overline{\ln(y)} \tag{20}$$

$$Var(u_1) = Var[\ln(x^*)] - b_1 Var[\ln(y^*)] \tag{21}$$

Once all the previous parameters are obtained, the joint distribution of ln(x*) and ln(y*) is calculated using 50,000 simulated households. Then, the simulated data is used to calculate the index of economic mobility and to construct a transition matrix.

Growth rate of per capita expenditures

The main advantage of Glewwe and Dang's method over Luttmer's method previously described is that it, by estimating the joint distribution of per capita expenditures in 2004 and 2006, allows us to calculate the estimated grow of per capita expenditures for each quintile corrected for measurement error. We then used the simulations obtained by applying Glewwe and Dang's as described in the previous section to calculate the growth rate in per capita expenditures corrected by measurement error. In order to ensure that the same households are compared, the growth rate in per capita expenditures is calculated using 2004 quintiles. In other words, the per capita expenditures of one particular quintile of households according to 2004 data are compared with per capita expenditures of the same group of households in 2006, regardless their position in the distribution of per capita expenditures in 2006.

IV. Data

The data used in this paper comes from the Peru´s National Household Surveys (***Encuesta Nacional de Hogares – ENAHO***), which has been implemented on a quarterly basis since 1997. These surveys are nationally representative and provide information on household demographics, education, health, employment, expenditure, income, and agricultural activities among others. Specifically, we use the surveys conducted in 2004, 2005, and 2006. The ENAHO surveyed 19,590 households in 2004, 20,577 in 2005, and 20,378 in 2006.

One of the advantages of these data sets is that they incorporate a nationally representative panel of 3,897 households interviewed once a year from 2004 to 2006. Each household included in the panel was interviewed in the same month each year. Thus, if a household was, for example, interviewed in April of 2004, it was also interviewed in April in 2005, and in April 2006. This allows us to assume that observed changes in household living conditions are not explained by season changes in expenditures or income.

Total expenditures in Peruvian Household Surveys are obtained by the aggregation of various groups of expenditures (i.e. food, transportation, housing, education, recreation, etc.). This information comes from the recall of expenditures obtained through household interviews. Most of the questions of these interviews are answered by the head of the household or the housewife. The average length for the whole interview is around three hours.

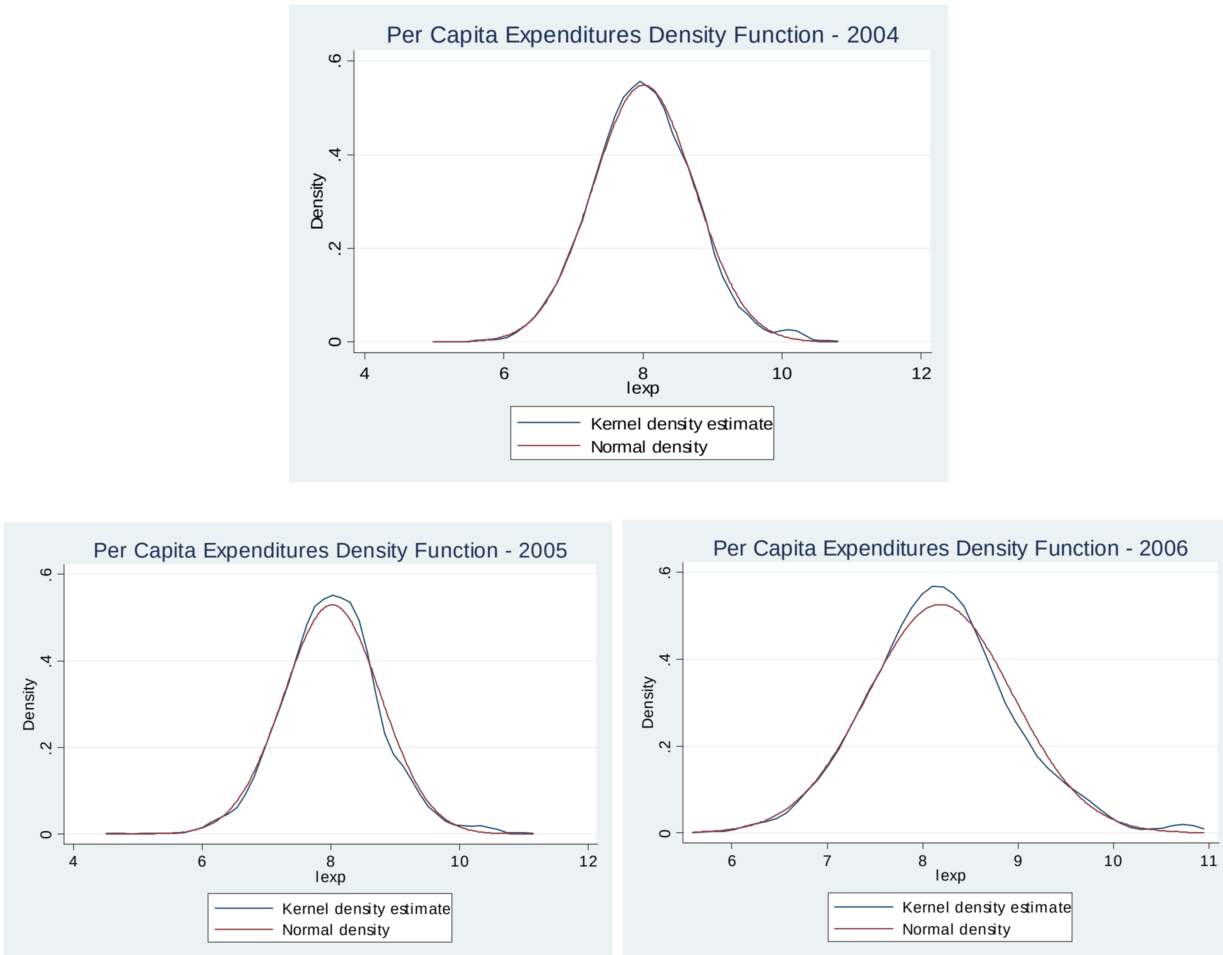
V. Results

Before presenting our results, we need to evaluate the assumption that the unobserved household expenditures follow a lognormal distribution, which is used to develop the simulations described in Section III. Assuming that measurement error follows a lognormal distribution, we would only need to check whether the observed household expenditures are also normally distributed[6]. Of course, it can be argued that the measurement error might not follow a lognormal distribution. However, we rely on the simulations provided by Glewwe (2007) which show that if we relax this assumption and use alternative distributions for measurement error (i.e. t-distribution or a gamma distribution), we will obtain very similar results for the mean of household expenditures.

The normality assumption is evaluated in Figure 1, where nonparametric kernel density estimates using the automatic (optimal) bandwidth for the logarithm of the observed per capita expenditures are displayed. Also, a Kurtosis-Skewness test was conducted to evaluate the log-normality assumption more formally. As we can observe in Figure 1, the observed household expenditures did not have a perfect fit to the normal distributions with the same mean and variance in any of the three years. Moreover, in all the cases the Kurtosis-Skewness test rejected the normality assumption. However, the normality assumption still provides a reasonably close fit, and it is not expected to distort the simulations.

[6] Recall that the sum of two normally distributed random variables has a normal distribution, and that the logarithm of the measured household expenditure is the sum of the logarithm of the true unobserved household expenditure and a random measurement error.

Figure 1: Density estimates of the logarithm of observed per capita expenditures



Estimates of measurement error

This section presents the impact of measurement error on the observed variance of the logarithm of per capita expenditures using Luttmer (2002)'s and Glewwe and Dang (2005)'s methods, as described in Section IV.[7] For the former, the variance of the logarithm of measurement error was estimated using each year of the data in hand. For the latter, the

[7] The parameters estimated using Glewwe and Dang's method and that will also be used in simulations are presented in Appendix 1.

variance of the measurement error was estimated only for 2004 and 2006, since those were the years used to estimate Equations 11 and 12. The results obtained are displayed in Table 1.

Table 1: Peru- Impact of Measurement Error on the Variance of the Logarithm of Per Capita Expenditures

	2004	2005	2006
Variance of observed per capita expenditures	0.53	0.57	0.58
Lutter (2002)'s method			
Variance of the log of per capita expenditures	0.47	0.50	0.51
Variance of the log of measurement error (lower bound)	0.06	0.07	0.06
Increased variance of log of per capita expenditures (%)	11.11	11.69	11.11
Glewwe and Dang (2005)'s method			
Variance of the log of per capita expenditures	0.46	-	0.50
Variance of the log of measurement error (lower bound)	0.07	-	0.08
Increased variance of log of per capita expenditures (%)	13.62	-	13.62

Table 1 shows that variance of the measurement error varies between 0.06 and 0.08, depending on the method used to its estimation and year. The results obtained by Luttmer's method suggest that it accounts for more than 11 percent of the observed variance of per capita expenditures. In contrast, under Glewwe and Dang's method it accounts for more than 13 percent of the observed variance of per capita expenditures. Since for Luttmer's method the estimated variance of the logarithm of the measurement error constitutes the lower bound of its real value; and for Glewwe and Dang's method the estimated variance of the logarithm of per capita expenditures represents the upper bound of its real value, all the estimates of the impact of measurement error on per capita expenditures should be considered conservative.

Simulations of economic mobility and transition matrices

Table 2 presents the estimates of economic mobility and the transition matrix using data that has not been corrected for measurement error between 2004 and 2006. The index for economic mobility, as defined in Equation 20, is equal to 0.240, which is relatively high considering that we are looking at changes over a period of two years. Between 2004 and 2006, only 45.8 percent of the households remain in the same quintile, 39.8 percent move by one quintile, and 14.4 percent move by two or more quintiles.

Table 2: Peru 2004/2006: Transition Matrix and Estimated Mobility in Per Capita Expenditure
Results Based on Observed Data

		2006 Quintile				
		1	2	3	4	5
2004 Quintile	1	12.8	4.9	1.7	0.5	0.1
	2	5.0	6.9	5.1	2.4	0.6
	3	1.4	5.0	6.5	5.4	1.7
	4	0.7	2.4	4.6	7.1	5.1
	5	0.1	0.8	2.1	4.6	12.5
Mobility index ($1-\rho[\ln(x_{2004}), \ln(x_{2006})]$):				0.240		

Calculations made based on observed per capita expenditures measured with error provide estimates of economic mobility that are biased upward (Glewwe, 2005). In this study, we provide two distinct approaches to address this problem. First, per capita expenditures in 2004 and changes in per capita expenditures are adjusted using Luttmer's (2002) method in order to calculate the per capita expenditures in 2006 corrected by measurement error. As described in Section IV, 50,000 simulations were generated to estimate economic mobility and create the transition matrices under Luttmer's method. Our results are displayed in Table 3.

As expected, economic mobility in Table 3 is lower than economic mobility based on observational data. In particular, the index of economic mobility is now 0.186, which implies that economic mobility in Table 2 was overestimated by about 24 percent. Also, Table 3 suggests that 47.9 percent of households remained in the same quintile over the two periods. This is around 5 percent higher than the correspondent percentage displayed in Table 2.

Table 3: Peru 2004/2006: Transition Matrix and Estimated Mobility in Per Capita Expenditure. Results Based on Observed Data Corrected for Measurement Error using Luttmer's method

		2006 Quintile				
		1	2	3	4	5
2004 Quintile	1	13.2	4.8	1.6	0.3	0.0
	2	4.9	7.5	5.2	2.1	0.3
	3	1.6	5.2	6.6	5.1	1.6
	4	0.3	2.2	5.1	7.5	5.0
	5	0.0	0.3	1.6	5.0	13.1
Mobility index ($1-\rho[\ln(x_{2004}), \ln(x_{2006})]$):				0.186		

An alternative method is to explicitly estimate that joint distribution using Glewwe and Dang's methodology, as discussed in Section IV. Table 4 shows the transition matrix and estimated mobility index using the upper bounds for ***Var[log(x*)]*** and ***Var[log(y*)]*** as shown in Appendix 1. The most striking result in Table 4 is that the estimated mobility index is only 0.120, which is half the value obtained using observed per capita expenditures. Also, the results suggest that 54.6 percent of Peruvian households stayed in the same quintile over the two years, which is almost 20 percent higher than the correspondent percentage when observed data is analyzed.

Table 4: Peru 2004/2006: Transition Matrix and Estimated Mobility in Per Capita Expenditure
Results Based on Simulated Data Corrected for Measurement Error using Glewwe and Dang's method
Using Estimated Upper Bounds for Var[log(x*)] and Var[log(y*)]

		2006 Quintile				
		1	2	3	4	5
2004 Quintile	1	14.5	4.5	0.9	0.1	0.0
	2	4.6	8.8	5.2	1.4	0.1
	3	0.8	5.3	7.8	5.2	0.9
	4	0.1	1.4	5.3	8.9	4.4
	5	0.0	0.0	0.8	4.5	14.6
Mobility index (1-ρ[ln(x_{2004}), ln(x_{2006})]):				0.120		

Results from Table 4 assumed that the estimated upper bounds of Var[log(x*)] and Var[log(y*)] are unbiased. However, if the measurement errors between 2004 and 2006 are correlated, that assumption is implausible. An alternative assumption is that the variance of the logarithm of per capita expenditures accounts for 80 percent of its observed value. The results after making this assumption are displayed in Table 5. In this case, the estimated mobility index is 0.127, which implies that measurement error in expenditure data accounts for 47 percent of the observed economic mobility. Results from Table 5 also suggest that 53.6 percent of Peruvian households stayed in the same quintile between 2004 and 2006.

Table 5: Peru 2004/2006: Transition Matrix and Estimated Mobility in Per Capita Expenditure
Results Based on Simulated Data Corrected for Measurement Error using Glewwe and Dang's method
Assuming Var[log(x*)]= 0.8Var[log(x)] and Var[log(y*)]=0.8Var[log(y)]

		2006 Quintile				
		1	2	3	4	5
	1	14.3	4.7	1.0	0.1	0.0
	2	4.6	8.6	5.3	1.5	0.1
2004 Quintile	3	1.1	5.1	7.6	5.3	1.0
	4	0.1	1.5	5.2	8.7	4.5
	5	0.0	0.1	0.9	4.4	14.5
Mobility index (1-ρ[ln(x_{2004}), ln(x_{2006})]):				0.127		

In Table 6 the variance of the logarithm of per capita expenditures is assumed to account for 70 percent of its observed value. It shows an estimated mobility index is 0.143, which suggests that the measurement error in expenditure data accounts for 41 percent of the observed economic mobility. Also, Table 6 indicates that 52.0 percent of Peruvian households stayed in the same quintile between 2004 and 2006.

Table 6: Peru 2004/2006: Transition Matrix and Estimated Mobility in Per Capita Expenditure
Results Based on Simulated Data Corrected for Measurement Error using Glewwe and Dang's method
Assuming Var[log(x*)]= 0.7Var[log(x)] and Var[log(y*)]=0.7Var[log(y)]

		2006 Quintile				
		1	2	3	4	5
	1	14.0	4.7	1.2	0.1	0.0
	2	4.7	8.4	5.0	1.7	0.2
2004 Quintile	3	1.1	5.1	7.3	5.3	1.1
	4	0.1	1.6	5.3	8.2	4.7
	5	0.0	0.1	1.2	4.7	14.0
Mobility index (1-ρ[ln(x_{2004}), ln(x_{2006})]):				0.143		

To summarize the latter results, estimates of suggest that the index of economic mobility varies from 0.12 to 0.18, which in turn accounts for between 24 and 50 percent of the economic mobility from observed data measured with error.

Simulations of the growth rate of per capita expenditures

In this section, the estimated growth rates of per capita expenditures by quintiles corrected for measurement error are presented. The data come from the simulation results generated using Glewwe and Dang's method to correct the per capita expenditures for measurement error. In this section, the same assumptions about the Var[log(x*)] and Var[log(y*)] are made as in the last section. In all cases the same households are compared based on the distribution of per capita expenditures in 2004. The results are displayed in Table 7.

Table 7: Peru 2004/2006: Estimated Annual Growth in Per Capita Expenditure by Quintiles

		Observed	Glewwe and Dang's method		
			Upper bounds	Var(y*)=0.8Var(y)	Var(y*)=0.7Var(y)
		(a)	(c)	(d)	(e)
2004 Quintile	1	24.1	15.0	15.6	16.6
	2	18.1	12.7	12.9	13.0
	3	15.5	11.0	11.2	11.0
	4	10.4	9.1	9.0	8.7
	5	4.5	6.4	6.1	4.5

The first conclusion to draw from Table 7 is that measurement error in per capita expenditures tends to underestimate its true growth rate for the poorest households. This is consistent with the finding in Glewwe (2007), which shows that when inequality increases over time while the extent of measurement error remains unchanged, measurement error leads to an overestimate of the growth of income (or expenditures) among the poorest households. Using the variance of the logarithm of per capita expenditures as an index of inequality, results from Table 1 suggests that the same logic can be applied in the Peruvian case, since inequality has increased while the variance of measurement errors in per capita expenditures have stayed more or less stable.

The results suggest that measurement error in per capita expenditure produces an overestimation of its growth rate that ranges from 46 up to 61 percent for the households in the first quintile, and from 39 up to 43 percent for the second quintile. For the third quintile the overestimation in the growth rate of per capita expenditures is of about 40 percent. For the fourth quintile, the overestimation varies between 14 and 19 percent. Finally, results for the fifth quintile suggest that the direction of the bias depends on the assumptions about the variances of the logarithm of per capita expenditures.

VI. Conclusions

This study represents an empirical exercise aimed to estimate the effects of measurement error in expenditure data using information from a Peruvian household panel for years 2004 to 2006. In particular, it intended to obtain unbiased estimators of economic mobility and the growth rate of per capita expenditures, after correcting for the effects of measurement error.

The results suggest that measurement error in expenditure data accounts for at least 11 to 13 percent of the observed variance of per capita expenditures. The study also shows that mobility indices estimated from observed data are biased upward, and that the magnitude of the bias ranges between 41 up to 50 percent. Finally, the results suggest that measurement error in per capita expenditure produces an overestimation of the growth rate of per capita expenditures that ranges from 46 to 61 percent for the households in the first quintile, and from 39 to 43 for the second quintile. For the third quintile the overestimation in the growth rate of per capita expenditures is of about 40 percent. For the fourth quintile, the overestimation varies between

14 and 19 percent. Finally, results for the fifth quintile suggest that the direction of the bias depends on the assumptions about the variances of the logarithm of per capita expenditures.

All these results reflect the importance of considering the effects of measurement error in household surveys when estimating economic indicators such as those of this study, or other indicators such as poverty or inequality indices. For instance, by knowing that economic mobility is not as high as observed data would indicate, focalization of poverty alleviation programs towards particular areas would not need to be revised as often compared to what it would be required if economic mobility were higher. Also, for evaluation of social programs that are designed to increase household income for a certain population group, similar methods may be applied to improve the estimated impacts of such programs.

Unfortunately, there are no previous studies to our knowledge that address this issue using data from Peru. We believe it is important that future research incorporate these considerations in order to provide more reliable evidence of Peru's economic performance. This in turn can improve the quality of the information used by policymakers to design and evaluate public policies and programs.

References

Antman, Francisca & David McKenzie. 2005. “Earnings Mobility and Measurement Error: A Pseudo-Panel Approach”, World Bank Policy Research Working Paper # 3745.

Battistin, Erich. 2002. “Errors in survey reports of consumption expenditures”. IFS Working Papers, W03/07. The Institute of Fiscal Studies.
http://www.ifs.org.uk/wps/wp0307.pdf

Bound, John and Alan Krueger. 1991. “The extent of measurement error in longitudinal earnings data: do two wrongs make a right?”. Journal of Labor Economics, 12, pp. 345–368.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. “Measurement error in survey data”, in: J. Heckman and E. Leamer (eds) Handbook of Econometrics, 5 (Amsterdam: Elsevier), pp. 3705–3843.

Deaton, Angus. 2000. “Consumption”. Chapter XVII in Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience, edited by Margaret Grosh and Paul Glewwe. The World Bank.
http://www.princeton.edu/~rpds/downloads/deaton_grosh_consumption.pdf

Deaton, Angus. 1997. “The Analysis of Household Surveys: A Microeconometric Approach to Development Policy”. (Baltimore: The Johns Hopkins University Press).

Duval Hernández, Robert, Gary Fields, Samuel Freije, María Laura Sánchez. 2006. "Income Mobility in Latin America". ILR Collection, Cornell University.
http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1011&context=workingpapers

Glewwe, Paul. 2007. "Measurement Error Bias in Estimates of Income and Income Growth among the Poor: Analytical Results and a Correction Formula". Economic Development and Cultural Change, 56 (1), pp. 163-190.

Glewwe, Paul. 2005. "How Much of Observed Economic Mobility Is Measurement Error? A Method to Reduce Measurement Error Bias, with an Application to Vietnam." Unpublished paper.
http://www.apec.umn.edu/faculty/pglewwe/documents/VNMOBIL3_05.pdf

Glewwe, Paul and Hai-Anh Hoang Dang. 2005. "Was Vietnam's Economic Growth in the 1990's Pro-Poor? An Analysis of Panel Data from Vietnam". Unpublished paper.
http://www.apec.umn.edu/faculty/pglewwe/documents/vnpropo4.pdf

Luttmer, Erzo F. P. 2002. "Measuring Economic Mobility and Inequality: Disentangling Real Events from Noisy Data." Unpublished paper, Kennedy School of Government, Harvard University.

Neter, John. 1970. "Measurement Errors in Reports of Consumer Expenditures". Journal of Marketing Research, 7(1), pp. 11-25.

Pischke, Jorn-Steffen. 1995. “Measurement Error and Earnings Dynamics: Some Estimates From PSID Validation Study”. Journal of Business and Economic Statistics 13(3), pp. 305-314.

Wooldridge, Jeffrey. 2002. “Econometric Analysis of Cross Section and Panel Data”. Massachusetts Institute of Technology.

Appendix 1

Estimates parameters proposed by Glewwe and Dang (2005) [1]

Parameter	Estimate	Method
$Var[\ln(x)]$	0.5287	From observed data
$Var[\ln(y)]$	0.5812	From observed data
$\overline{\ln(x)}$	8.0105	From observed data
$\overline{\ln(y)}$	8.1658	From observed data
b_2	0.7972	OLS regression from observed data
b_2^*	0.9229	IV regression given by Equation 12
b_1	0.7251	OLS regression from observed data
b_1^*	0.8394	From proportional variance assumption given by Equation 13
a_1^*	1.1565	Calculated using Equation 21
$Var[u_1]$	0.1029	Calculated using Equation 22
Upper bound for $Var[\ln(x^{\bullet})]$	0.4567	Calculated using Equation 13
Upper bound for $Var[\ln(y^*)]$	0.5021	Calculated using Equation 13

(1) All estimates based on 3,897 household with complete data in 2004 and 2006