Measurement Error in Access to Markets *

Javier Escobal
GRADE

Sonia Laszlo
McGill University

jescobal@grade.org.pe
sonia.laszlo@mcgill.ca

March 22, 2005

*Magdalena Benza provided assistance to process GPS referenced data and to construct 'true' travel time variables. We thank seminar participants at the department of Economics and the Social Statistics seminar at McGill University, GRADE, and conference participants at the 44e Congrès de la Société Canadienne des Sciences Économiques (Québec) and at the 75 Years of Development Research International Colloquium (Cornell) for their insightful comments

Abstract

Studies in the microeconometric literature increasingly utilize distance to or time to reach markets or social services as determinants of economic issues. These studies typically use self-reported measures from survey data, often characterized by non-classical measurement error. This paper is the first validation study of access to markets data. New and unique data from Peru allow comparison of self-reported variables with scientifically calculated variables. We investigate the determinants of the deviation between imputed and self-reported data and show that it is non-classical and dependent on observable socio-economic variables. Our results suggest that studies using self-reported measures of access may be estimating biased effects.

1 Introduction

Access to markets and social infrastructure is often considered an important determinant of economic behavior in developing countries, particularly in rural areas. For instance, distances to schools are inversely related to educational attainment (e.g. Glewwe and Jacoby (1994)) and thus can have a negative impact on human capital acquisition. Rosero-Bixby (2004) finds a negative effect between distance to a health facility and the probability of choosing that health facility. Similarly, distance to family planning facilities are thought to have an impact on contraceptive use (though Mroz et al. (1999) find little effect). Geographic proximity to markets affects the types of economic activity that people engage in, affecting both wages, participation and the distribution of income (e.g. De Janvry and Sadoulet (2001) and Escobal (2001)). Distances, times to reach markets and social infrastructure and geographic proximity are used in a multitude of other studies to predict policy relevant outcomes.[1]

To be useful for policy-making, empirical analyses of access to markets depends on the availability of data to measure and approximate access to markets. Most often, access to markets data come from self-reported answers from survey data. However, we know that survey data in general is prone to measurement error. Several literatures have documented substantial measurement error in self-reported data for developed countries (particularly in the reporting of earnings or health status). Measurement error is just as, if not more, likely to exist in data sets from developing countries, and most empirical studies using such data sets acknowledge such limitations. Surprisingly, while statistical effort to validate self-reported data from developed countries has increased over the last decade, little effort has been placed to do so in the context of developing country data sets. In this paper,

[1] For example, Seeth et al. (1998) use the time it takes to get from the place of residence to the plot as an explanatory variable in a time allocation framework. Zaal and Oostendorp (2002) use travel time to market in a study of small-scale agriculture in Kenya and find that it has a significant effect on the probability that the plot is terraced. Swain (2002) uses distance from village to the nearest concrete road in rural India to explain demand and access to formal and informal credit. Jacoby (2000) finds that longer travel times lead to lower plot values and lower agricultural wages.

we are able to validate self-reported data on access to markets. This is to the best of our knowledge the first validation study of access to markets data and the first to validate data from a developing country.

The empirical labor economics literature has recently given a great deal of attention to the accuracy with which earnings data is reported in labor market income surveys such as the Current Population Survey (CPS) and Panel Study of Income Dynamics (PSID) data for the US (Bound and Krueger (1991), Bound et al. (1994), Pischke (1995), Brownstone and Valleta (1996), Bollinger (1998) and Hyslop and Imbens (2001)). Through-out this literature, authors consistently find evidence of measurement error and of its attenuation bias. Card (1996) validates union status by investigating employer and employee records in the US CPS. The health economics literature has also appealed to validation studies to assess the validity of self-reported health data. Using administrative and self-reported health data from Canada, Baker, Stabile and Deri (2005) provide an excellent example of how measurement error in self-reported health assessments can lead to serious inference problems (particularly since respondents report both false negatives and false positives). Other examples of validation studies in the health and health economics literature include Norton et al. (2003) and Biemer and Wiesen (2002).

Surprisingly, there is little, if any, work seeking the degree to which such measurement error exists in developing countries and how this may bias the results using household surveys from such areas. Given our own experiences doing fieldwork in the Peruvian Andes, and based on reports from our surveyors in earlier surveys, we have often observed significant underreporting of distance and time by local populations. This paper is, to the best of our knowledge, the first to assess measurement error in data on access to (distance to or time to reach) markets. Many studies in the development literature use distance to social services, to markets, to agricultural plots and so on to explain a number of policy relevant questions. However, given the lack of validation of these types of data, little is known about whether the estimated parameters are biased. If access to markets is measured with error, its effects may be biased towards zero if measurement error is classical. Furthermore, if the error in measurement is correlated with observable socio-economic characteristics and thus non-classical, it is unclear what the estimated effects are truly picking up. Identifying those variables that are highly correlated with the true travel time variable but not with the measurement error will provide insights about the kind of instrumental variables needed to correct for potential measurement error biases.

Using a unique, Global Positioning System (GPS) validated, data set from rural Peru, we determine the degree to which rural households in this developing country make errors in self-reported travel time variables. The next section of this paper addresses some of the econometric issues surrounding measurement error, in particular in the case that it is correlated with other explanatory variables. In section 3, we describe our unique data. Section 4 provides a description and an analysis of measurement error. Section 5 concludes. The appendix provides details about our GPS validation method.

2 Measurement Error

2.1 Theory

Suppose that we are interested in estimating the effect of socio-economic characteristics and access to markets on an individual's outcome. Access to markets, proxied reasonably well by the time it takes for the individual to reach the market, is potentially measured with error. This section outlines some of the econometric problems estimating this relationship by OLS when one such explanatory variable is measured with error. Without losing generality of the conclusions and, to make matters simple, the model we illustrate here is parsimonious: we consider only one socio-economic variable that, along with the dependent variable, is precisely measured.

We begin with a very simple econometric model, with a classical measurement error in one independent variable. We then augment this model with non-classical measurement error where the measurement error in one explanatory variable is correlated with another explanatory variable. To do so, we illustrate the biases in OLS estimates that result from a simple linear regression model with two only explanatory variables. Consider the following relationship:

$$y^* = b_1 x_1^* + b_2 x_2^* + e \tag{1}$$

where y^* is the dependent variable, x_1^* and x_2^* are the explanatory variables, b_1 and b_2 are the parameters that we wish to estimate and e is the residual. The asterisks denote true variables, not necessarily observed. However, while x_2^* is not observed, we do observe x_2 such that:

$$x_2 = x_2^* + u \tag{2}$$

where u is our measurement error. Thus the relationship between the observed dependent variable and explanatory variables is:

$$y^* = b_1 x_1 + b_2 x_2^* + e - b_2 u \tag{3}$$

We assume that $e \sim iid(0, \sigma_e^2)$, $e \sim iid(0, \sigma_u^2)$, $cov(x_i, e) = 0\ \forall i \in 1, 2$, $cov(e, u) = 0$ and $u \perp x_2^*$, such that $cov(x_2^*, u) = 0$. This last assumption implies that $cov(x_2, u) = \sigma_u^2$. We make the following two assumptions to describe the case of classical measurement error:

$$cov(x_1, u) = 0 \tag{4}$$

$$cov(x_1, x_2) = 0 \tag{5}$$

If assumptions (4) and (5) hold, then it can easily be shown that the OLS estimate of b_2 is inconsistent and has the following bias:

$$plim(\hat{b_2} - b_2) = -b_2 \frac{\sigma_u^2}{\sigma_{x_2^*}^2 + \sigma_u^2} \tag{6}$$

where $\sigma_{x_2^*}$ is the variance of x_2^*. Thus under the classical errors-in-variables model (where (4) and (5) hold), if access to markets is measured with error, its estimated effect on the outcome y is biased towards zero. Such measurement error does not affect the OLS estimate of the effect of the other independent variables.

There are however numerous reasons to believe that the measurement error is correlated with socio-economic variables. Under these circumstances, assumption (4) will be violated. Consider instead the following assumption:

$$cov(x_1, u) = \sigma_{x_1,u} \neq 0 \tag{7}$$

As a result, a correlated measurement error violates (5).

Proposition 1 *Let μ_{x_1} and μ_{x_2} denote the means of x_1 and x_2, respectively. Since $cov(x_1, u) = \sigma x_1 u \neq 0$ and x_1 and x_2 are not independent, it can easily be shown that $cov(x_1, x_2) = \sigma_{x_1 x_2} = E[x_1 x_2] - \mu_{x_1}\mu_{x_2} \neq 0$.*

The correlations between x_1 and u or x_2 have significant implications for the consistency of the OLS parameters, b_1 and b_2. Under these circumstances, the OLS estimate of the effect of the precisely measured explanatory variable will also be biased and the expressions for the biases become complicated. It can be shown that the biases are as follows:

$$plim(\hat{b_1} - b_1) = -b_2 \frac{\sigma^2_{x_2}\sigma_{x_1 u} - \sigma^2_u(\sigma_{x_1 x_2} + \mu_{x_1}\mu_{x_2})}{\sigma^2_{x_1}\sigma^2_{x_2} - (\sigma_{x_1 x_2} + \mu_{x_1}\mu_{x_2})^2} \tag{8}$$

$$plim(\hat{b_2} - b_2) = -b_2 \frac{\sigma^2_{x_2}\sigma^2_u - \sigma_{x_1 u}(\sigma_{x_1 x_2} + \mu_{x_1}\mu_{x_2})}{\sigma^2_{x_1}\sigma^2_{x_2} - (\sigma_{x_1 x_2} + \mu_{x_1}\mu_{x_2})^2} \tag{9}$$

Note that by setting $\sigma_{x_1 x_2} = \sigma_{x_1 u} = 0$ and letting x_1 and x_2 be independent, the expressions in (8) and (9) reduce to the results from the OLS estimates under the classical measurement error model. Under this more general case where $\sigma_{x_1 x_2} \neq 0$ and $\sigma_{x_1 u} \neq 0$, the signs and relative magnitudes of the biases are ambiguous and rely on the covariances between the explanatory variables and the measurement error. Knowing σ^2_u and $\sigma_{x_1 u}$ will allow us to identify the direction and magnitude of the bias.

2.2 Empirical Strategy

This paper aims to characterize measurement error in access to markets. After computing the measurement error as $u_i = x_{2i} - x^*_{2i}$, we obtain and qualify the variance-covariance matrix for the vector of variables $\psi = (\mathbf{X'_{1i}}, x_{2i}, x^*_{2i}, u)'$. This variance-covariance matrix will yield some key moments influencing the biases in the OLS estimates as depicted in the model above. In addition, we estimate whether the measurement error in access to markets is indeed correlated with other socio-economic characteristics:

$$u_i = \mathbf{X'_{1i}}\alpha_{\mathbf{x}} + \zeta_i \tag{10}$$

Finally, to gauge the degree of bias under the non-classical measurement error model, we attempt to evaluate the ratios in (8) and (9). For a number of potential x_2, we calculate the following ratios:

$$\theta_1 = \frac{\sigma_{x_2}^2\sigma_{x_1u} - \sigma_u^2(\sigma_{x_1x_2} + \mu_{x_1}\mu_{x_2})}{\sigma_{x_1}^2\sigma_{x_2}^2 - (\sigma_{x_1x_2} + \mu_{x_1}\mu_{x_2})^2} \tag{11}$$

$$\theta_2 = \frac{\sigma_{x_2}^2\sigma_u^2 - \sigma_{x_1u}(\sigma_{x_1x_2} + \mu_{x_1}\mu_{x_2})}{\sigma_{x_1}^2\sigma_{x_2}^2 - (\sigma_{x_1x_2} + \mu_{x_1}\mu_{x_2})^2} \tag{12}$$

3 The Data

3.1 The Survey

The data that we utilize come from a two-round household survey conducted by GRADE. The first round was conducted between and July and August 2003. Its goal was to understand the determinants of the demand for technical assistance by Peruvian agricultural producers. The second round in December 2003 revisited a sub sample of the producers interviewed in the first round with the intent to understand the links between policy and access to markets for these producers. This second round survey collected geographic information and perceptions of time, among other variables. The survey asked respondents to answer a number of questions relating to their agricultural production, access to technical help, access to markets, along with other socio-economic characteristics (such as age, education, religion, ethnicity and family composition of the household head). It surveyed three broadly different geographic and ecological zones. From the Selva, 270 coffee producers were sampled in 10 districts from the Amazonas, Cajamarca and San Martín departments. From the Sierra, 260 potato producers were sampled in 23 districts in the department of Junín. From the Costa, 202 rice producers were sampled in 8 districts in the department of Lambayeque. It should be noted that the samples here are non-representative, either at a national or even regional level.[2]

To gauge access to markets, the survey includes two modules of particular interest for the present study. First, a module in the first round asks each respondent to self-report the time (in minutes) and transportation method that it takes him or her to reach his/her most important plot, furthest plot, the nearest public telephone, the agrarian agency, the medical centre, the nearest primary school, the nearest secondary school, the nearest interprovincial bus stop, the nearest market and the nearest credit provider. Second, a module in the second round on the perception of time includes the following five questions:[3]

1. Q: How to you tell time? A: By looking at own watch; By asking someone else; By looking at the position of the sun; Doesn't case about the time; Other

2. Q: Do you or someone in your dwelling have a watch?

3. Q: Is there a clock or alarm clock in your dwelling

4. Q: How long does it take you to walk from you dwelling to the center of the nearest populated center?[4]

5. Q: Method of travel and travel time to get from your dwelling to the nearest district capital?

In rural communities, such as those in Peru, one expects time to have a different meaning than it does in more developed areas. Without delving into some of the more ethnological or anthropological reasons, we recognize that cultural relevance is a potentially driving component of measurement error. While the dimension of the problematic

[2]Since these samples were constructed in order to capture how small farmers react to new market opportunities, the areas chosen need not be representative of each of these three regions. For example, for the Sierra, the study sampled small farms in one of the most dynamic valleys in the Andes. Farmers in these areas are not relatively well integrated in input and output markets.

[3]This information is only available for the coffee and potato samples. In the sample of potato producers, only 177 households answered these questions.

[4]A populated center is equivalent to a village.

should not be overlooked, it reinforces the importance of this study, and only highlights the data collection problems in culturally different environments. If this is indeed the case, then relying too much on self-reported measures of access based on time may lead to erroneous predictions.

To validate the data, in the second round of the survey true walking distance were identified for one out of ten respondents chosen randomly. Surveyors walked with the respondent from their dwelling to the center of the nearest populated center, following the same route and at the same pace as the respondent (the instructions were clear to that matter). The surveyor then recorded how long it took for the trajectory. In addition, the surveyor recorded latitude, longitude and altitude, along with the trajectory using a Global Positioning System (GPS) device.[5] This data was then used to calculate the true time from the dwelling to the center for the entire sample accounting for the type and quality of terrain, its slope and navigability. This measure is our 'true measure' for x_2^*. The appendix describes the methodology employed to calculate this measure.

Ideally, we would consider the respondent's reply to question 4 "How long does it take you to walk from your dwelling to the center of the nearest populated center?" to be the 'self-reported measure' for x_2. However, respondents often misinterpreted this question and instead answered how long it took them to walk to the nearest populated center, *other than their own.* While this is a form of measurement in its own right, we will concentrate on another measure. As discussed above, the respondents were asked how long it took them to reach the nearest primary school, secondary school, health center, etc. Typically, populated centers have a primary school and the primary school is most often located in the center (Plaza de Armas), sometimes adjacent to the church.[6] Most individuals walk to the primary school rather than taking any other form of transportation.[7] Thus, we feel

[5]We do not claim to be the only ones using GPS data for access. Rosero-Bixby (2004) for instance uses GIS data to geo-reference health facilities in Costa Rica. However, their distances measure the 'as the crow flies', whereas our data is able to geo-reference actual trajectories.

[6]We verify this against a school census conducted by the Ministry of Education in 2000. We keep only observations for which there's a primary school in the populated center.

[7]We include only observations where individuals walk to the nearest primary school. Observations in which individuals report the time to the nearest primary school taking a bus or a taxi are dropped.

confident that self-reported time to the nearest primary school is adequately picking up time to the center of the populated center. In what follows, we describe the measurement error derived by taking the difference between the self-reported and the true times.

3.2 Descriptive Statistics

We begin by providing the descriptive statistics of the data in Table 1. The data set is separated into three samples: Selva, Costa and Sierra. Comparing the self-reported and the true time to the center of the nearest populated center (the first two rows), we see that respondents consistently under-report the time it takes them to reach the center. The third row provides the difference between the self-reported and the true measure - this is our measure of measurement error. In addition, we observe large regional differences: in the Sierra, average measurement error is at its lowest (in magnitude) with -2.173 minutes; the average measurement error in the Selva is -6.255 minutes; and the average in the Costa is highest at about -11.609 minutes. We also report the absolute value of the measurement error, as some of the analysis in what follows utilizes the degree to which there is an error (rather than the direction). The regional patterns described for measurement error are more pronounced when looking at the absolute value. In the fifth row, we also present the self-reported time to the nearest center, taken from the module on the perception of time. It is evident that the respondents are making an error in their interpretation of the center, rather than in the time it takes them to reach it: the average reported time in the Selva is about 58 minutes, while the average true time is just over 12 minutes!

Household heads are oldest in the Costa, and youngest in the Selva. In most cases the household heads are male. They are least educated in the Costa, and most educated in the Sierra. Over 80% of households report someone owning a wrist watch, though the proportion is highest among Sierra households. This is probably a result of the large income differential between these regions. Average household income in the Sierra is twice that in the Selva. Land ownership, however, is greatest in the Selva. Finally, the household background variables, such as household size and the proportion of children in the household not attending school, are relatively stable across regions.

Table 1 provides additional information about the perception of time. Particularly, for the Selva and Sierra samples, we observe how people tell time. Sixty percent of households in the Selva and 73.6% in the Sierra tell time by looking at their watches. This differential is undoubtedly a reflection in the watch ownership differential between the two samples. It is particularly interesting to note that, in the Sierra, 14.5% of households tell time by looking at the sun's position. Few (less than 1%) report not caring about the time.

Table 2 provides cross-tabulations between the self-reported data and our true measure of time by sample in order to complete the description of our data.[8] We split each variable into two categories: below and above median value for each variable.[9] Panel A , B and C provide the cross-tabulations for the coffee, rice and potato samples, respectively. In the three panels, we see that the majority of observations lie on the main diagonals. For instance, 45.42% coffee producers self-reported the time as below the median when the true time is also below the median, and 20.61% reported above the median where the true measure is also above the median. Confirming our descriptive statistics above, we note that respondents are more likely to under-report the time as the off-diagonal item to the left is larger than the off-diagonal element to the right. These patterns, while not as stark, are also observed for the other two samples. One conclusion to be drawn from this table is that there is a great deal of noise in these data.

4 Analysis

Figures 1 to 3 present the kernel densities for the self-reported and the true times for the Selva, Costa and Sierra, respectively. These figures confirm that respondents are

[8]This representation of measurement error is inspired by Card's 1996 paper on measurement error in union status.

[9]Because of low variation in the self-reported measure, the number of observations per cell is different as we move along each column - there are numerous observations at the median. Ideally, we would have broken this up into a larger number of quintiles, but this low variation in the self-reported measure would lead to very small cells.

consistently under-reporting the time it takes them to reach the nearest populated center. The patterns in the Selva (figure 1) are particularly interesting - respondents seem to be clumping' their self-reported time around integers such as 10, 15, 20 and 30 minutes. This pattern is not as stark in the other two samples, and we provide no interpretation of this difference in patterns other than cultural differences across the three regions.

Figure 4 describes the distribution of the measurement error for the three regions. We see that the distribution of the measurement error is different across different samples and that it is skewed to the left (reflecting that respondents under-report the time it takes them to reach the center). Measurement error is greatest for the rice sample and lowest for the coffee sample.[10].

Tables 3, 4 and 5 provide the correlation matrices for the variables for the coffee, rice and potato samples. In the first column, we observe the correlation between the measurement error (u) and the observables: age, sex and education of the household head, household income and size, land holdings, children not in school and (except in the case of rice) watch ownership and how the household head tells time. The patterns of correlations are very different across the three samples. For the coffee sample, it appears that education, household size, and whether the household head tells time by looking at a watch or asking others are correlated with the measurement error. However, our interpretation varies depending on the correlation with the different components of this error. Education and household size are correlated with the true time, not with the self-reported time: more educated households are more likely to live in or closer to the center of the town. Larger households tend to live further from the center of town. However, being more educated or having a larger household does not influence the self-reported distance in this sample. Conversely, telling time by looking at the watch is significantly negatively correlated to the self-reported time and not with the true time. For the rice sample, altitude is the only variable to be significantly correlated with the measurement error. Here, altitude is correlated with both components of the measurement error: households

[10]We leave explanations for the reasons why measurement error is different across these samples for future research

in more elevated areas are more isolated and the true time to the center of the town is greater. Meanwhile, households in more elevated areas tend to report shorter times to reach the center. In the potato sample, altitude and living on the land used for agricultural production are correlated with the true time to the center of the town. None of the variables are correlated with the self-reported time in this sample.

Based on the evidence provided by the correlation matrices in these three samples, measurement error in self-reported time to the nearest populated center seems uncorrelated with the majority of the observable socio-economic characteristics. This bodes well for most studies using self-reported time as an explanatory variable: measurement error is not likely to cause biases in inferences beyond the standard classical measurement error bias. This is certainly the case for the sample of Sierra potato producers. However, there is a relatively strong correlation between the measurement error and the manner in which time is read for the coffee and potato samples, and the altitude of the dwelling in the rice sample.

Telling time by looking at the watch reduces the error in the self-reported time to the populated center in the coffee sample. Naturally, this is not a surprising result, and is entirely consistent with intuition. Furthermore, studies would rarely include watch ownership in a regression that would also include self-reported time to the nearest populated center. However, the correlation between watch ownership and the measurement error could still contaminate the estimates, via its correlation with household income, a commonly employed socio-economic characteristic. Indeed, the correlation between watch ownership and household income is significantly positive (see the 7^{th} column in table 3). While income is not directly correlated with the measurement error, it may very well be indirectly correlated via watch ownership. We now turn to some parametric results. We estimate equation (10) for each of the three samples separately. We present the main results in table 6 where the OLS estimates of equation (10) for the coffee (columns (1) to (3)), rice (column (4)) and potato samples (columns (4) to (7)). We use the absolute value of measurement error to isolate the determinants of the magnitudes, rather than the direction, of the error. For coffee and potatoes, the data allows us to test several specifications making use of the perception of time. Columns (1) and (5) include watch ownership as an explanatory variable, while columns (2) and (6) instead include the method of telling time.[11]

The results in table 6 depict a relatively weak linear fit of the data. Only in the rice sample can we reject the statistical significance of the regression. In addition, the Z-test for the normality of the regression residuals is strongly rejected in all cases. Nonetheless, some of the observable socio-economic characteristics seem to be significantly correlated with measurement error. For the coffee and the potato samples, the education of the household head is negatively related with the magnitude of the measurement error: more educated individuals make smaller errors. The altitude of the dwelling is positively related to measurement error in the rice sample only: at higher altitudes, dwellings and populated centers tend to be more isolated and distances tend to be larger, thus increasing the scope for error.[12]

Finally, tables 7 to 9 present the imputed biases calculated for each of the three samples solving equations (11) and (12). It is clear from these tables that the signs and magnitudes of these biases are 'all over the place'. For example, a correlation between household income and measurement error in the coffee sample would lead to an attenuation bias in a regression of household income and distance to the nearest populated center on an outcome variable y, in both coefficients (not just in the one measured with error). Generally, the imputed biases in all samples are far from 1, suggesting large biases. It is particularly interesting to note that the degree of bias (i.e. how far away θ_1 and θ_2 are from 1) is most often largest for the variables measured with error than it is for its covariate (that is, $|\theta_1| > |\theta_2|$ in most cases). Furthermore, these biases can be tremendously large: they can increase the magnitudes of the estimated OLS coefficients by a factor of as much as 27,477 in the case of 'children not in school' in the rice sample. About half of the imputed biases are statistically significant, particularly with respect to the age, gender and education of the household head, household size and the altitude of the dwelling.

[11] The omitted category here is telling time by looking at one's watch.

[12] Altitude is only very weakly significant in one of the potato specifications.

5 Conclusion

This paper describes and analyses measurement error in a common explanatory variable in the field of economic development: access to the nearest market (proxied here as the time it takes to walk to the nearest populated centre). Classical measurement error, uncorrelated with other explanatory variables, leads to an attenuation bias in the regression coefficient on the variable measured with error. In this paper, we have qualified and quantified the bias in the event that measurement error is non-classical. In these cases, the biases can be tremendously large, with no particular general pattern, and can lead to overestimates as well as underestimates of the regression coefficients on the variable measured with error, as well as its co-variates.

Using a unique and validated data set from coffee, rice and potato producing areas in rural Peru, this paper is the first to validate access to markets data. This paper is also the first, that we are aware of, to conduct a validation study utilizing data from a developing country. We are able to reject classical measurement error in some cases: the degree to which survey respondents answer how long it takes them to reach the nearest population center is positively correlated with the altitude of their dwelling, the presence of children not attending school, and negatively related to watch ownership (which is itself positively correlated with income). Classical measurement error is generally rejected in the coffee (Selva) and rice (Costa) samples. However, the results show that, in the case of the potato sample (Sierra), measurement error is only correlated with altitude and not the other observable socio-economic characteristics.

The implication of this study is that researchers using self-reported access to markets or public infrastructure measured by time may need to address the likelihood that the responses are reported with a great deal of error. In addition to finding that this error is correlated with some other socio-economic characteristics in the case of two of the three Peruvian samples (coffee and rice), we also found that respondents often misunderstood the question 'how long does it take you to walk to the nearest populated center', another type of measurement error altogether.

Appendix: Measuring Accessibility

We define 'the true' measurement of accessibility as the time it takes for an individual to walk from his or her dwelling to the center (plaza de armas) of the nearest town. Considering that that we only registered the actual path and timing for a sample of producers (about 1 of every 10), it was necessary to calculate the access times for all others producers. This was done taking into account the geographical coordinates of their dwelling and of the center of the town as well as the characteristics of the terrain that those producers must travel. The first two pieces of information were collected as part of the survey while the third comes from secondary sources mentioned below.

The available information on the characteristics of the terrain between the dwelling and the center of town includes the map of roads (engineered motorized roads, and non-engineered dirt roads), available from the Ministry of Transportation and Communication as well as the complete cartography of the National Geographical Institute (for the case of the tracks, trails and footpaths, which are typically defined as non-motorized rural roads). The map of rivers and gulches comes also from the cartography of the National Geographical Institute. Finally, the elevation map used comes from the Digital Elevation Model obtained by the Shuttle Radar Topography Mission of NASA. This cartography was transformed to a raster format. This type of Geographic Information System (GIS) format assumes that the cartography has been ordered in grid form. Each grid represents, through the cells that it encompasses, the values that best describe the characteristics of that space. For example, the map of roads in grid form indicates in each cell if there is a road or not as well as the type of road. Overlapping of these grids creates a friction surface map that summarizes all land characteristics of the terrain, using information of slopes, natural barriers, roads and floor usage. Distance and travel times can be calculated from this grid as the least effort required to move from a cell to contiguous cell.

All cartography used was based on geographical coordinates using the WGS 84 datum. However, for the purpose of the analysis, the cartography was projected to the Universal Transverse Mercator (UTM) plane coordinate system; specifically into area 18, which is the predominant area for Peru. Due to the fact that UTM coordinate system works with metric and not degree measurements, calculations of distances and travel times are greatly facilitated using this kind of projection when we focus on small areas, as is the case in this study.

The first step in creating this friction surface is to have all the spatially related variables in a grid format with comparable resolutions. In this case we used a scale-resolution of 90 meters by 90 meters for each cell, because this it is the maximum resolution that can be obtained from the Digital Elevation Model (DEM). Using this DEM, a map of slopes was created which constitutes the starting point for the calculation of a friction surface. The map of slopes was used to calculate the walking speeds through the terrain were producers travel. Following Tobler (1993), these walking speeds were applied to all roads and were also applied areas not covered by the road network, according to the following formula:

$$Speed = \lambda\gamma[\alpha\gamma \exp(\beta\gamma|s + \gamma|)] \tag{13}$$

Here s represents the slope, λ represents a penalty in the speed when people decide to walk off-road (for example crossing a plot), α, β and γ are calculated in such a way that we can obtain from (13) true speeds gathered in the field.[13] Using estimations done by Tobler(1993), and field verification done within this study we used the following parameter values for equation 1: $\alpha = 6$; $\beta = -3.5$; $\gamma = 0.05$; and $\lambda = 0.6$ for walking off road. These parameters reflect the true speeds that were gathered in different areas of the studies walking on-road and off-road in flat as well as in hilly terrain. Figure A1 depicts how walking velocities are affected by the slope of the terrain depending on whether the producer is traveling on-road or off-road.

After calculating the speed at which a person moves from each cell of the grid to a neighboring cell, is necessary to combine this information with the physical and natural barriers that potentially can prevent a person from moving into contiguous cells in the grid. To incorporate this feature, we included rivers and gulches as natural barriers. We also took into account the existence of bridges that would overcome these barriers.

[13]Note that if people walk in road rather than off-road, $\lambda\gamma = 1$.

The walking speeds calculated from the map of slopes were combined with the map of roads and with the map of barriers (rivers and gulches) to obtain the friction map. This friction surface is again a grid were each cell has a value related to the time it takes to travel through that particular cell. This time varies according to the terrain characteristics.

After the friction surface was built, we applied a Cost Distance algorithm following ESRI (2002) which adds cumulatively the values of each cell following a critical path starting from the towns of interest around each one of the points of origin. The resulting map is a new grid of cells where the values no longer correspond to the time to walk through that cell, but to the accumulated time needed to walk to that cell from the point of origin. In this new grid it is possible to identify the required time to arrive from any point from the map to the nearest town.

Validation: After optimal trajectories were obtained given the estimated walking speeds and the physical and natural barriers faced by the producer we compared the results of this GIS modeling exercise and simulated optimal trajectories with the true travel times and trajectories of that sub-sample of producers that were monitored in field. Figure 2, for example, shows the trajectory of a coffee producer located in the district of Alonso of Alvarado, province of Lamas, in the region of San Martin (in the Peruvian Amazon). The GIS-simulated and effective trajectories show a similar pattern. However in a few cases in the validation phase, the calculated routes did not coincide exactly with the observed routes. In most of theses cases we found that this was so because rural road network was not fully up to date, so we opt for maintaining the calculated speeds from which the friction map and accessibility measure were calculated.

References

[1] Baker, Michael, Mark Stabile and Catherine Deri. "What Do Self-Reported, Objective, Measures of Health Measure?" *Journal of Human Resources* Vol. 39 (2004).

[2] Biemer, Paul P. and Christopher Wiesen. "Measurement Error Evaluation of Self-Reported Drug Use: A Latent Class Analysis of the US National Household Survey on Drug Abuse." *Journal of the Royal Statistical Society A* Vol. 165 (Part 1) (2002): 97-119.

[3] Bolan, Marc. "The Mobility Experience and Neighborhood Attachment." Demography Vol. 34 (2) (May 1997): 225-37. Bollinger, Christopher R. "Measurement Error in the Current Population Survery: A Non-Parametric Look." *Journal of Labor Economics* Vol. 16 (3) (1998): 576-94.

[4] Borooah, Vani K. "Births, Infants and Children: An Econometric Portrait of Women and Children in India." *Development and Change* Vol. 34 (1) (2003): 67-103.

[5] Bound, John. "Self-Reported versus Objective Measures of Health in Retirement Models." *Journal of Human Resources* Vol. 26 (1) (1990): 106-38.

[6] Bound, John and Alan B. Krueger. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics* Vol. 9 (1) (1991): 1-24.

[7] Bound, John, Charles Brown, Greg J. Duncan and Willard L. Rodgers. "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data." *Journal of Labor Economics* Vol. 12 (3) (1994): 345-68.

[8] Brener, Nancy D., Tim McManus, Deborah A. Galuska, Richard Lowry, Howell Wecshler. "Reliability and Validity of Self-reported Height and Weight Among High School Students." *Journal of Adolescent Health* Vol. 32 (2003): 281-7.

[9] Brouwer, Inge D., Jan C. Hoorweg and Marti Van Liere. "When Households Run out of Fuel: Responses of Rural Households to Decreasing Fuelwood Availability, Ntcheu District, Malawi." *World Development* Vol. 25 (2) (1997): 255-66.

[10] Brownstone, David and Robert G. Valletta. "Modeling Earnings Measurement Error: A Multiple Imputation Approach." *Review of Economics and Statistics* Vol. 78 (4) (1996): 705-17.

[11] Büchel, Felix and Maarten van Ham. "Overeducation, Regional Labor Markets, and Spatial Flexibility." *Journal of Urban Economics* Vol. 53 (2003): 482-93.

[12] Butler, J.S., Richard V. Burkhauser, Jean M. Mitchell and Theodore P. Pincus. "Measurement Error in Self-Reported Health Variables." *Review of Economics and Statistics* Vol. 69 (4) (1987): 644-50.

[13] Card, David. "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis." *Econometrica* Vol. 64 (4), July (1996): 957-79.

[14] Crossley, Thomas F. and Steven Kennedy. "The Reliability of Self-Assessed Health Status." *Journal of Health Economics* Vol. 21 (2002): 643-58.

[15] De Janvry, Alain and Elisabeth Sadoulet. "Income Strategies Among Rural Households in Mexico: the Role of Off-Farm Activities." *World Development* Vol. 29 (3) (2001): 467-80.

[16] Duncan, Greg J. and Daniel H. Hill. "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data." *Journal of Labor Economics* Vol. 3 (4) (1985): 508-32.

[17] Dwyer, Debra Sabatini and Olivia S. Mitchell. "Health Problems as Determinants of Retirement: Are Self-Rated Measures Endogenous?" *Journal of Health Economics* Vol. 18 (1999): 173-93.

[18] Escobal, Javier. "The Determinants of Nonfarm Income Diversification in Rural Peru." *World Development* Vol 29 (3) (2001): 497-508.

[19] ESRI. "ArcView 3.3 Cost Distance Modeling Discussion: help facility." Environmental Systems Research Institute (ESRI) Inc., 380 New York St., Redlands, CA 92373, http://www.esri.com, (2002).

[20] Glewwe, Paul and Hanan Jacoby. "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana." *Journal of Human Resources* Vol 29(3) (Summer 1994): 843-64.

[21] Hyslop, Dean R. and Guido Imbens. "Bias From Classical and Other Forms of Measurement Error." *Journal of Business and Economic Statistics* Vol. 19(4): 475-81.

[22] Jacoby, Hanan. "Access to Markets and the Benefits of Rural Roads." *Economic Journal* Vol. 110 (July 2000): 713-37.

[23] Linde-Rarh, Martin. "Property Rights and Deforestation: the Choice of Fuelwood Source in Rural Viet Nam." *Land Economics* Vol. 79 (2) (May 2003): 217-34.

[24] Miller, Harvey J. "Measuring Space-Time Accessibility Networks: Basic Theory and Computational Procedures." *Geographical Analysis* Vol. 31 (1999): 187-212.

[25] Mroz, Thomas A., Kenneth A. Bollen, Ilene S. Speizer and Dominic J. Mancini. "Quality, Accessibility, and Contraceptive Use in Rural Tanzania." *Demography* Vol. 36 (1) (February 1999): 23-40.

[26] Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa and Jose Luis Evia. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* Vol. 16 (2) (2002): 241-74.

[27] Norton, Edward C., Richard C. Lindrooth and Susan T. Ennett. "How Measures of Perception From Survey Data Lead to Inconsistent Regression Results: Evidence from Adolescent and Peer Substance Use." *Health Economics* Vol. 12 (2003): 139-48.

[28] Pischke, Jorn-Steffen. “Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study.” *Journal of Business and Economic Statistics* Vol. 15 (3) (July 1995): 305-14.

[29] Rosero-Bixby, Luis. “Spatial Access to Health Care in Costa Rica and its Equity: a GIS-Based Study.” *Social Science and Medicine* Vol 58 (2004): 1271-84.

[30] Seeth, Harm Tho, Sergei Chachanov and Alexander Surinov. “Russian Poverty: Muddling Through Economic Transition with Garden Plots.” *World Development* Vol. 26 (9) (1998): 1611-23.

[31] Sjoquist, David L. “Spatial Mismatch and Social Acceptability.” *Journal of Urban Economics* Vol. 50 (2001): 474-90.

[32] Swain, Ranjula Bali. “Credit Rationing in Rural India.” *Journal of Economic Development* Vol. 27 (2) (December 2002): 1-20.

[33] Tobler, Waldo, 1993 “Three presentations on Spatial Analysis: ‘Non-Isotropic Modeling’, ‘Speculations on the Geometry of Geography’, ‘Global Geographical Analysis’.” NCGIA Technical Report TR-93-1, Santa Barbara, 39pp, (1993).

[34] Zaal, Fred and Remco H. Oostendorp. “Explaining a Miracle: Intensification and the Transition Towards Sustainable Small-Scale Agriculture in Dryland Machakos and Kitui Districts, Kenya.” *World Development* Vol. 30 (7) (2002): 1271-87.

Table 1

Descriptive Statistics						
	Selva - Coffee Producers		Costa - Rice Producers		Sierra - Potato Producers	
	N=262		N=166		N=159	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Self-reported time to the nearest populated center‡ (minutes)	6.721	8.139	14.964	10.988	6.899	4.747
True time (minutes)	12.976	17.729	26.573	25.364	9.072	8.899
Measurement error	-6.255	17.299	-11.609	28.407	-2.173	9.159
Absolute value of measurement error	10.026	15.415	18.615	24.370	6.012	7.230
Self-reported time to the nearest populated center† (minutes)	57.748	49.308	--	--	21.465	29.959
Age of household head	41.874	13.559	53.572	15.536	48.717	13.619
Household head is male	0.969	0.172	0.952	0.215	0.950	0.219
Years of schooling of the household head	6.225	3.436	5.241	3.439	10.648	3.535
Someone in household has a watch	0.817	0.388	--	--	0.931	0.255
Household size	4.729	2.108	4.904	2.189	4.692	1.779
Household income (Soles per year)	10,004.140	12,526.300	12,833.360	24,582.280	19,407.390	25,275.760
Land ownership (ha)	7.280	25.126	3.729	6.002	2.605	5.200
Children not attending school	0.019	0.137	0.006	0.078	0.013	0.112
Altitude of dwelling (meters above sea level)	1,248.366	282.926	45.861	11.982	3,393.629	122.220
Lives on land used for agriculture and livestock production	0.221	0.416	0.590	0.493	0.208	0.407
Method of telling time						
Tells time by looking at watch	0.603	0.490	--	--	0.736	0.442
Tells time by asking others	0.076	0.266	--	--	0.050	0.219
Tells time by looking at the sun's position	0.050	0.218	--	--	0.145	0.353
Doesn't care about the time	0.008	0.087	--	--	0.006	0.079
Tells time using other method	0.263	0.441	--	--	0.063	0.244

Note: * Samples include only observations where people reported walking to the nearest primary school, and in populated centers that report having a primary school in the 2000 school census. ‡ is the measure of self-reported time based on the time to walk to the nearest primary school. This is the measure used to calculate measurement error. † is the measure of self-reported time which is picking up the time to the nearest populated center, *other than their own.*

Table 2

Cross Tabulations of Self-Reported versus True Time, by sample

A. Coffee		Self-Reported Time	
		Below Median	Above Median
True Time	Below Median	119	12
		(45.42%)	(4.58%)
	Above Median	77	54
		(29.39%)	(20.61%)

B. Rice		Self-Reported Time	
True Time		Below Median	Above Median
	Below Median	64	19
		(38.55%)	(11.45%)
	Above Median	52	31
		(31.33%)	(18.67%)

C. Potatoes		Self-Reported Time	
		1st Quartile	2nd Quartile
True Time		Below Median	Above Median
	Below Median	56	24
		(35.22%)	(15.09%)
	Above Median	40	39
		(25.16%)	(24.53%)

Note: due to relatively low variation in the self-reported measure, the 'below median' and 'above median' columns do no represent a 50/50 split.

Table 3

Correlation matrix - Coffee Sample (Selva)

	u	Self-reported time	True time	Age	Male	Educa-tion	Income	Hhsize size	Land holdings	Kids not in school	Altitude of dwelling	Lives on land used for agriculture and livestock	Watch	Tell time (Other)	Tell time (Watch)	Tell time (Ask)	Tell time (Sun)	Tell time (don't care)
u	1																	
Self-reported time	0.182*** (0.003)	1																
True time	-0.892*** (0.000)	0.282*** (0.000)	1															
Age	-0.054 (0.389)	-0.004 (0.946)	0.050 (0.418)	1														
Male	-0.030 (0.632)	-0.012 (0.852)	0.024 (0.703)	-0.072 (0.245)	1													
Education	0.137** (0.026)	-0.014 (0.826)	-0.140** (0.023)	-0.419*** (0.000)	0.063 (0.307)	1												
Income	0.002 (0.979)	-0.076 (0.222)	-0.036 (0.558)	0.052 (0.406)	0.041 (0.509)	0.155** (0.012)	1											
Hhsize size	-0.143*** (0.020)	0.016 (0.801)	0.147** (0.017)	0.1465** (0.018)	0.051 (0.412)	-0.156** (0.011)	0.041 (0.505)	1										
Land holdings	0.006 (0.919)	-0.051 (0.414)	-0.029 (0.635)	0.039 (0.533)	0.036 (0.564)	0.055 (0.375)	0.060 (0.337)	-0.061 (0.328)	1									
Kids not in school	-0.098 (0.113)	0.115* (0.064)	0.148** (0.016)	0.078 (0.211)	0.025 (0.690)	-0.066 (0.287)	-0.034 (0.586)	0.270*** (0.000)	0.001 (0.992)	1								
Altitude of dwelling	0.071 (0.254)	0.109* (0.080)	-0.019 (0.758)	-0.016 (0.797)	0.052 (0.403)	-0.016 (0.793)	-0.100 (0.106)	0.015 (0.813)	-0.061 (0.324)	0.065 (0.292)	1							
Lives on land used for agriculture and livestock	-0.062 (0.319)	0.330*** (0.000)	0.212*** (0.001)	0.082 (0.187)	0.041 (0.507)	-0.107* (0.083)	-0.015 (0.806)	0.047 (0.450)	-0.022 (0.722)	0.060 (0.333)	-0.028 (0.656)	1						
Watch	-0.103 (0.095)	-0.151** (0.014)	0.032 (0.611)	0.017 (0.788)	0.088 (0.156)	0.028 (0.649)	0.131** (0.034)	0.066 (0.290)	0.055 (0.375)	-0.006 (0.922)	-0.043 (0.493)	-0.104* (0.093)	1					
Tell time (Other)	0.093 (0.133)	0.139** (0.025)	-0.027 (0.662)	0.010 (0.871)	-0.045 (0.469)	-0.042 (0.501)	-0.148** (0.017)	-0.014 (0.827)	-0.056 (0.371)	0.107* (0.085)	0.248*** (0.000)	0.120* (0.053)	-0.434*** (0.000)	1				
Tell time (Watch)	-0.134** (0.030)	-0.154** (0.013)	0.060 (0.331)	-0.053 (0.392)	0.083 (0.182)	0.081 (0.194)	0.197** (0.001)	0.014 (0.820)	0.053 (0.393)	-0.058 (0.351)	-0.182*** (0.003)	-0.112* (0.070)	0.584*** (0.000)	-0.737*** (0.000)	1			
Tell time (Ask)	0.103* (0.095)	0.045 (0.466)	-0.080 (0.197)	-0.004 (0.953)	-0.116* (0.061)	-0.082 (0.187)	-0.053 (0.395)	-0.059 (0.345)	-0.002 (0.973)	-0.040 (0.518)	-0.073 (0.237)	-0.015 (0.812)	-0.161*** (0.009)	-0.172*** (0.005)	-0.354*** (0.000)	1		
Tell time (Sun)	0.026 (0.671)	0.017 (0.791)	-0.018 (0.770)	0.062 (0.319)	0.041 (0.513)	0.011 (0.864)	-0.072 (0.247)	0.063 (0.311)	-0.008 (0.904)	-0.032 (0.608)	-0.004 (0.951)	0.048 (0.444)	-0.210*** (0.001)	-0.137** (0.027)	-0.282*** (0.000)	-0.066 (0.290)	1	
Tell time (don't care)	-0.099 (0.111)	-0.019 (0.765)	0.088 (0.156)	0.105* (0.091)	0.016 (0.802)	-0.019 (0.765)	-0.021 (0.741)	0.011 (0.856)	0.008 (0.901)	-0.012 (0.844)	0.003 (0.964)	-0.047 (0.451)	-0.072 (0.247)	-0.052 (0.398)	-0.108* (0.081)	-0.025 (0.685)	-0.020 (0.747)	1

N=262

Table 4
Correlation Matrix - Rice Sample (Costa)

	u	Self-reported time	True Time	Age	Male	Educa-tion	Income	Hhld Size	Land holdings	Kids not in school	Altitude of dwelling	Lives on land used for agriculture and livestock
u	1											
Self-reported time	0.456*** (0.000)	1										
True time	-0.923*** (0.000)	-0.0769 (0.325)	1									
Age	0.128 (0.101)	0.093 (0.232)	-0.103 (0.188)	1								
Male	-0.032 (0.680)	-0.032 (0.687)	0.023 (0.773)	-0.053 (0.494)	1							
Education	-0.085 (0.279)	0.007 (0.929)	0.098 (0.211)	-0.543*** (0.000)	-0.042 (0.595)	1						
Income	0.070 (0.370)	0.110 (0.159)	-0.031 (0.693)	0.073 (0.348)	0.021 (0.787)	0.012 (0.878)	1					
Hhld Size	0.026 (0.739)	0.229*** (0.003)	0.070 (0.370)	0.155** (0.046)	0.080 (0.304)	-0.104 (0.183)	0.049 (0.528)	1				
Land holdings	0.095 (0.225)	0.106 (0.172)	-0.060 (0.442)	0.193** (0.013)	0.019 (0.806)	-0.077 (0.323)	0.921*** (0.000)	0.029 (0.711)	1			
Kids not in school	0.073 (0.352)	0.107 (0.171)	-0.035 (0.653)	-0.058 (0.457)	0.018 (0.823)	-0.074 (0.346)	-0.017 (0.826)	0.039 (0.617)	-0.029 (0.711)	1		
Altitude of dwelling	-0.460*** (0.000)	-0.276*** (0.000)	0.396*** (0.000)	0.097 (0.216)	-0.092 (0.238)	0.085 (0.274)	-0.072 (0.354)	-0.072 (0.357)	-0.022 (0.777)	-0.025 (0.748)	1	
Lives on land used for agriculture and livestock	-0.101 (0.196)	0.019 (0.813)	0.121 (0.120)	0.171** (0.028)	-0.130* (0.094)	0.016 (0.841)	0.089 (0.252)	0.076 (0.334)	0.135* (0.083)	-0.094 (0.231)	-0.086 (0.273)	1

N=179

Table 5

Correlation matrix - Potatoes Sample (Sierra)

	u	Self-reported time	True time	Age	Male	Educa-tion	Income	Hhsize size	Land holdings	Kids not in school	Altitude of dwelling	Lives on land used for agriculture and livestock	Watch	Tell time (Other)	Tell time (Watch)	Tell time (Ask)	Tell time (Sun)	Tell time (don't care)
u	1																	
Self-reported time	0.233***	1																
	(0.000)																	
True time	-0.866***	0.285***	1															
	(0.000)	(0.000)																
Age	0.039	-0.013	-0.045	1														
	(0.552)	(0.846)	(0.492)															
Male	0.068	-0.097	-0.117	-0.072	1													
	(0.302)	(0.137)	(0.074)	(0.272)														
Education	0.069	-0.099	-0.119	-0.229***	0.186***	1												
	(0.295)	(0.131)	(0.070)	(0.000)	(0.004)													
Income	0.023	0.016	-0.015	-0.027	-0.060	0.259***	1											
	(0.727)	(0.812)	(0.825)	(0.677)	(0.360)	(0.000)												
Hhsize size	-0.016	-0.086	-0.029	-0.127*	0.135**	-0.080	-0.032	1										
	(0.808)	(0.189)	(0.664)	(0.052)	(0.038)	(0.220)	(0.624)											
Land holdings	0.077	-0.020	-0.086	0.122*	0.054	0.088	0.516***	-0.082	1									
	(0.241)	(0.762)	(0.189)	(0.062)	(0.406)	(0.177)	(0.000)	(0.209)										
Kids not in school	-0.032	-0.062	0.000	-0.027	0.022	0.047	-0.017	0.170***	-0.005	1								
	(0.626)	(0.346)	(0.996)	(0.679)	(0.732)	(0.471)	(0.801)	(0.009)	(0.944)									
Altitude of dwelling	0.209***	0.092	-0.158**	-0.023	0.048	-0.107	-0.097	0.081	-0.047	-0.009	1							
	(0.001)	(0.159)	(0.015)	(0.732)	(0.467)	(0.101)	(0.139)	(0.219)	(0.471)	(0.888)								
Lives on land used for	-0.209***	0.063	0.238***	0.102	0.066	-0.150**	-0.078	0.009	0.009	-0.044	0.016	1						
	(0.001)	(0.338)	(0.000)	(0.118)	(0.311)	(0.021)	(0.232)	(0.892)	(0.887)	(0.504)	(0.811)							
Watch	-0.024	-0.058	-0.006	-0.033	0.164**	0.163**	0.118	0.023	0.054	0.031	-0.032	-0.044	1					
	(0.764)	(0.466)	(0.937)	(0.679)	(0.039)	(0.041)	(0.138)	(0.778)	(0.503)	(0.700)	(0.693)	(0.583)						
Tell time (Other)	0.111*	0.013	-0.103	0.072	-0.041	-0.104	-0.082	0.005	-0.052	-0.020	-0.058	0.009	-0.338***	1				
	(0.090)	(0.838)	(0.117)	(0.269)	(0.530)	(0.111)	(0.210)	(0.938)	(0.429)	(0.766)	(0.373)	(0.888)	(0.000)					
Tell time (Watch)	-0.009	0.019	0.019	0.032	0.055	0.072	0.084	0.007	-0.002	0.093	-0.051	-0.009	0.455***	-0.210***	1			
	(0.888)	(0.774)	(0.775)	(0.622)	(0.403)	(0.272)	(0.202)	(0.919)	(0.979)	(0.155)	(0.435)	(0.891)	(0.000)	(0.001)				
Tell time (Ask)	0.014	-0.068	-0.048	0.036	-0.057	-0.042	-0.075	-0.173***	-0.055	-0.017	0.040	-0.028	-0.277***	-0.040	-0.187***	1		
	(0.837)	(0.299)	(0.461)	(0.579)	(0.383)	(0.524)	(0.253)	(0.008)	(0.401)	(0.791)	(0.540)	(0.666)	(0.000)	(0.546)	(0.004)			
Tell time (Sun)	-0.037	0.058	0.066	-0.065	0.017	-0.042	-0.036	0.029	0.006	-0.031	0.066	0.178***	-0.1697**	-0.069	-0.328***	-0.062	1	
	(0.578)	(0.378)	(0.316)	(0.323)	(0.795)	(0.524)	(0.583)	(0.659)	(0.933)	(0.642)	(0.312)	(0.006)	(0.033)	(0.289)	(0.000)	(0.345)		
Tell time (don't care)	0.007	-0.024	-0.019	-0.041	0.016	0.006	0.002	-0.102	0.003	-0.006	-0.058	-0.031	0.022	-0.014	-0.065	-0.012	-0.022	1
	(0.917)	(0.718)	(0.773)	(0.530)	(0.809)	(0.927)	(0.971)	(0.119)	(0.961)	(0.926)	(0.378)	(0.637)	(0.786)	(0.834)	(0.320)	(0.852)	(0.743)	

N=167

Table 6

Dependent Variable: $|u|$=| self-reported - true measure|							
	Coffee (N=262)			**Rice (N=166)**	**Potatoes (N=157)**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Age of household head	-0.040	-0.046	-0.039	-0.373*	-0.047	-0.046	-0.048
	(0.078)	(0.079)	(0.078)	(0.139)	(0.044)	(0.045)	(0.044)
Household head is male	3.891	3.996	4.334	3.810	0.561	0.680	1.011
	(5.566)	(5.591)	(5.549)	(8.038)	(2.811)	(2.786)	(2.760)
Education of household head	-0.575*	-0.575*	-0.570*	-0.318	-0.437**	-0.448**	-0.428**
	(0.311)	(0.313)	(0.311)	(0.621)	(0.191)	(0.192)	(0.191)
Household size	0.664	0.712	0.691	0.961	-0.181	-0.273	-0.183
	(0.480)	(0.481)	(0.479)	(0.799)	(0.345)	(0.355)	(0.345)
Land ownership (ha)	-0.005	-0.004	-0.004	-0.302	-0.129	-0.135	-0.133
	(0.038)	(0.038)	(0.038)	(0.346)	(0.125)	(0.125)	(0.125)
Children not in school	5.704	5.773	5.557	-6.511	0.430	0.506	0.519
	(7.207)	(7.247)	(7.207)	(22.218)	(5.296)	(5.306)	(5.290)
Log total household income	1.379	1.237	1.537	4.574	0.888	0.822	0.995
	(1.250)	(1.285)	(1.241)	(2.872)	(0.763)	(0.773)	(0.753)
Altitude of dwelling (meters above sea level)	-0.004	-0.004	-0.004	0.946***	-0.008	-0.008*	-0.008
	(0.003)	(0.003)	(0.003)	(0.149)	(0.005)	(0.005)	(0.005)
Have clock/watch	2.529				2.044		
	(2.487)				(2.358)		
Method of telling time ‡							
Tells time using other method		-1.533				-3.561	
		(2.342)				(2.438)	
Tells time by asking others		-1.504				-2.670	
		(3.721)				(2.795)	
Tells time by looking at the sun		-4.749				0.077	
		(4.511)				(1.678)	
Doesn't care about the time		17.655				-7.235	
		(10.974)				(7.389)	
Constant	-1.360	2.184	-1.204	-51.992*	28.709	34.383*	29.349
	(13.125)	(13.608)	(13.125)	(28.029)	(18.884)	(19.049)	(18.854)
R^2	0.048	0.060	0.044	0.231	0.058	0.077	0.053
Adj R^2	0.014	0.015	0.014	0.192	0.000	-0.0000	0.002
F stat	1.40	1.32	1.45	5.89***	1.00	1.00	1.03
Z-test for normality of residuals	9.199***	9.069***	9.209***	4.578***	7.704***	7.624***	7.714***

Notes: Standard errors in brackets. ***, **, * are significant at 1%, 5% and 10% respectively. ‡ omitted category is "tells time by looking at watch".

Table 7

Imputed Biases: Coffee Sample (Selva)

			μ_{x1}	σ^2_{x1}	σ_{x1u}	σ_{x1x2}	θ_1	θ_2
	$\sigma^2_{x_2}$	66.24						
	μ_{x2}	6.72						
	σ^2_u	299.26						
x_1=	Age		41.87	183.86	-12.55	-0.47	**1.27**	**-0.35**
							(0.22)	**(0.18)**
	Male		0.97	0.03	-0.09	-0.02	**48.44**	**-492.13**
							(8.11)	**(120.63)**
	Education		6.23	11.81	8.16	-0.38	**12.67**	**-20.80**
							(3.78)	**(9.76)**
	Household Size		4.73	4.44	-5.22	0.27	**13.56**	**-27.27**
							(2.74)	**(9.08)**
	Land holdings (ha)		7.28	631.34	2.75	-10.35	-0.28	0.49
							(19.57)	(7.19)
	Children not in school		0.02	0.02	-0.23	0.13	-78.08	16,811.39
							(48.41)	(14,891.95)
	Household income		10,004.14	156,908,191.69	352.05	-7,726.54	0.00	0.00
							(0.05)	(2.24)
	Altitude of dwelling		1,248.37	80,046.90	345.81	249.95	**0.04**	**0.04**
							(0.01)	**(0.03)**
	Watch ownership		0.82	0.15	-0.69	-0.48	101.83	-1,305.67
							(1,870.20)	(5294.55)
	Tells time (other)		0.26	0.19	0.71	0.50	-81.52	**2,557.11**
							(69.74)	**(743.83)**
	Tells time (watch)		0.60	0.24	-1.14	-0.61	-270.76	4,858.72
							(5,538.82)	(91,141.17)
	Tells time (ask)		0.08	0.07	0.48	0.10	**-35.08**	**4,593.84**
							(17.74)	**(1335.74)**
	Tells time (sun)		0.05	0.05	0.10	0.03	-33.94	**6,598.64**
							(19.05)	**(2,451.02)**
	Tells time (don’t care)		0.01	0.01	-0.15	-0.01	-42.48	39,468.21
							(23.24)	(24,502.43)

N=262. Bootstrapped standard errors in brackets (1000 replications).

Table 8

Imputed Biases: Rice Sample (Costa)								
			μ_{x1}	σ^2_{x1}	σ_{x1u}	σ_{x1x2}	θ_1	θ_2
	$\sigma^2_{x_2}$	120.74						
	μ_{x2}	14.96						
	σ^2_u	806.97						
x_1=	Age		53.57	241.37	56.39	15.93	**1.02**	**-0.08**
							(0.15)	**(0.05)**
	Male		0.95	0.05	-0.20	-0.07	**58.70**	**-499.24**
							(8.82)	**(122.78)**
	Education		5.24	11.83	-8.25	0.26	**13.54**	**-20.59**
							(2.69)	**(6.73)**
	Household Size		4.90	4.79	1.62	5.51	**11.24**	**-17.24**
							(1.78)	**(4.54)**
	Land holdings (ha)		3.73	36.02	16.15	7.02	-120.86	239.07
							(294.95)	(2387.57)
	Children not in school		0.01	0.01	0.16	0.09	**-182.73**	140,292.97
							(50.03)	(40,177.80)
	Altitude		45.86	143.56	-156.62	-36.29	**1.34**	**-0.49**
							(0.21)	**(0.10)**
	Household income		12,833.36	604,288,490.00	48,924.90	29,696.10	-0.01	-0.46
							(0.02)	(1.60)

N=166. Bootstrapped standard errors in brackets (1000 replications).

Table 9

Imputed Biases: Potato Sample (Sierra)								
			μ_{x1}	σ^2_{x1}	σ_{x1u}	σ_{x1x2}	θ_1	θ_2
	$\sigma^2_{x_2}$	22.53						
	μ_{x2}	6.90						
	σ^2_u	83.89						
x_1=	Age		48.72	185.48	12.49	-0.25	**0.26**	**0.02**
							(0.06)	**(0.02)**
	Male		0.95	0.05	0.14	-0.17	**13.42**	**-47.62**
							(3.33)	**(14.62)**
	Education		10.65	12.50	1.67	-1.84	**1.23**	**-0.37**
							(0.29)	**(0.11)**
	Household Size		4.69	3.16	-0.12	-0.78	**2.86**	**-2.04**
							(0.70)	**(0.59)**
	Land holdings (ha)		2.61	27.04	4.65	0.01	-4.90	6.31
							(161.41)	(123.87)
	Children not in school		0.01	0.01	-0.04	-0.04	-16.58	6,757.45
							(4.67)	(4521.28)
	Household income		19,407.39	638,864,043.58	6,002.64	7,600.39	0.00	0.15
							(0.02)	(10.10)
	Altitude of dwelling		3,393.63	14,937.73	283.52	96.06	**0.00**	**0.01**
							(0.00)	**(0.00)**
	Watch ownership		0.93	0.07	-0.06	-0.07	**13.74**	**-48.64**
							(3.34)	**(14.95)**
	Tells time (other)		0.06	0.06	0.31	0.01	**-26.76**	**1,654.94**
							(12.88)	**(738.47)**
	Tells time (watch)		0.74	0.20	-0.17	-0.01	**20.18**	**-89.09**
							(5.51)	**(31.51)**
	Tells time (ask)		0.05	0.05	0.03	-0.10	**-19.88**	1,855.31
							(8.33)	(1,061.53)
	Tells time (sun)		0.14	0.12	-0.17	0.11	-61.33	1,197.93
							(235.58)	(746.00)
	Tells time (don't care)		0.01	0.01	0.01	-0.01	**-17.84**	13,431.72
							(6.41)	(4,556.27)

N=159. Bootstrapped standard errors in brackets (1000 replications).

Figure 1

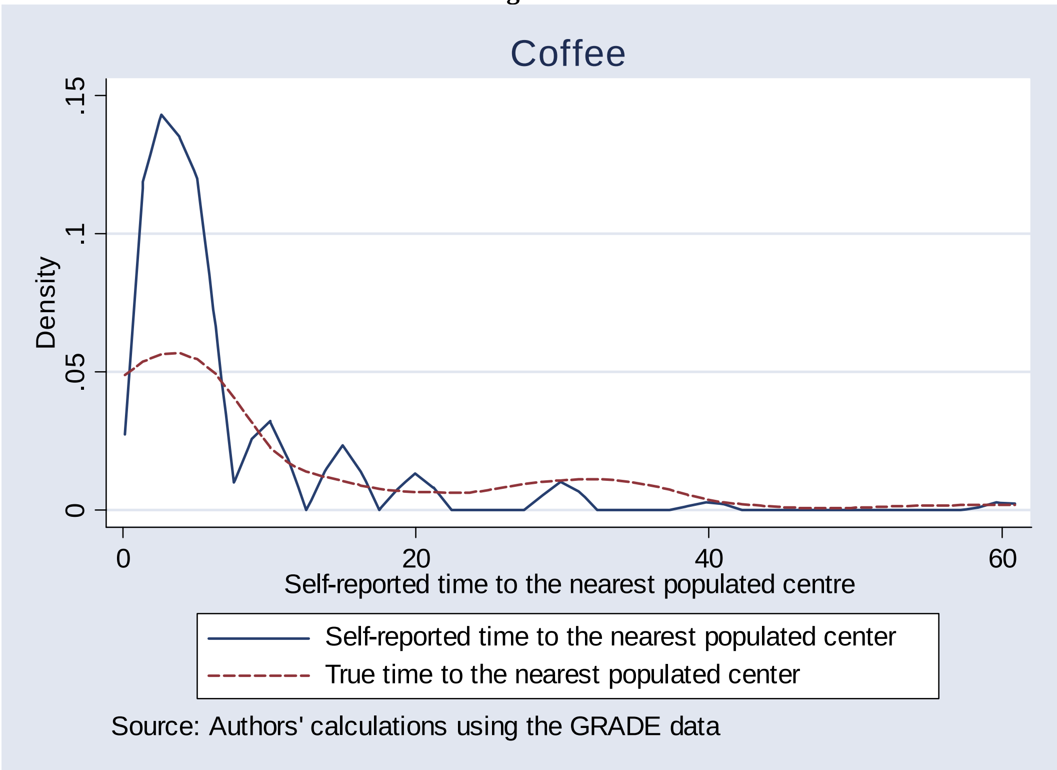

Figure 2

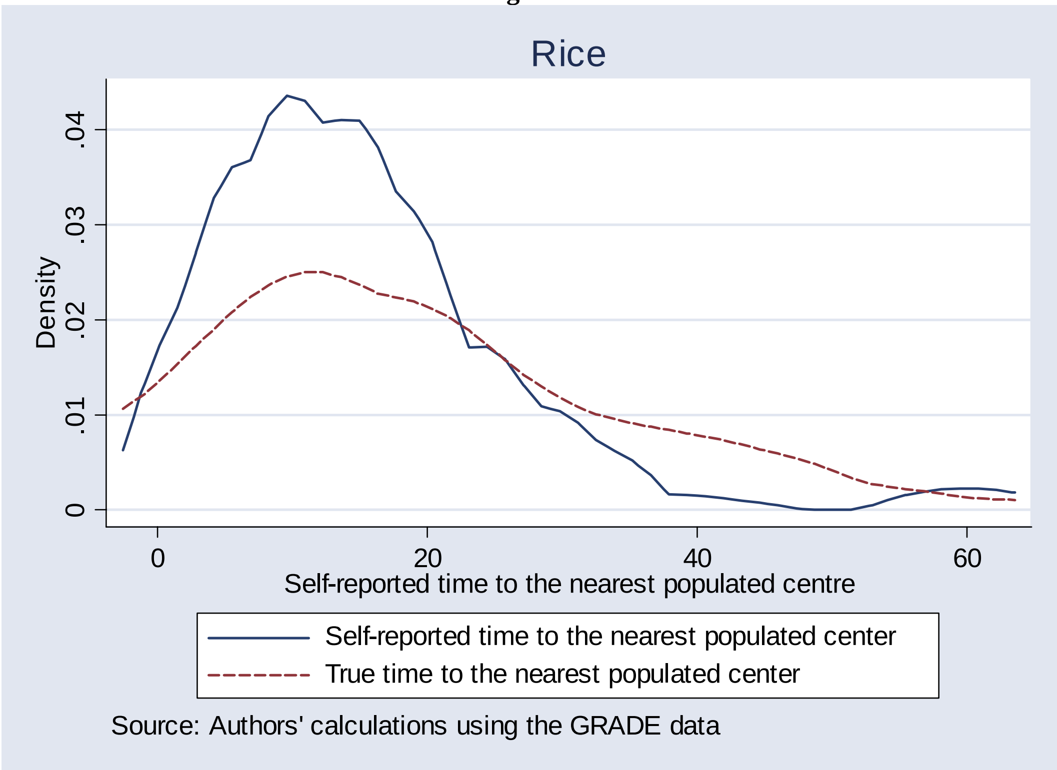

Figure 3

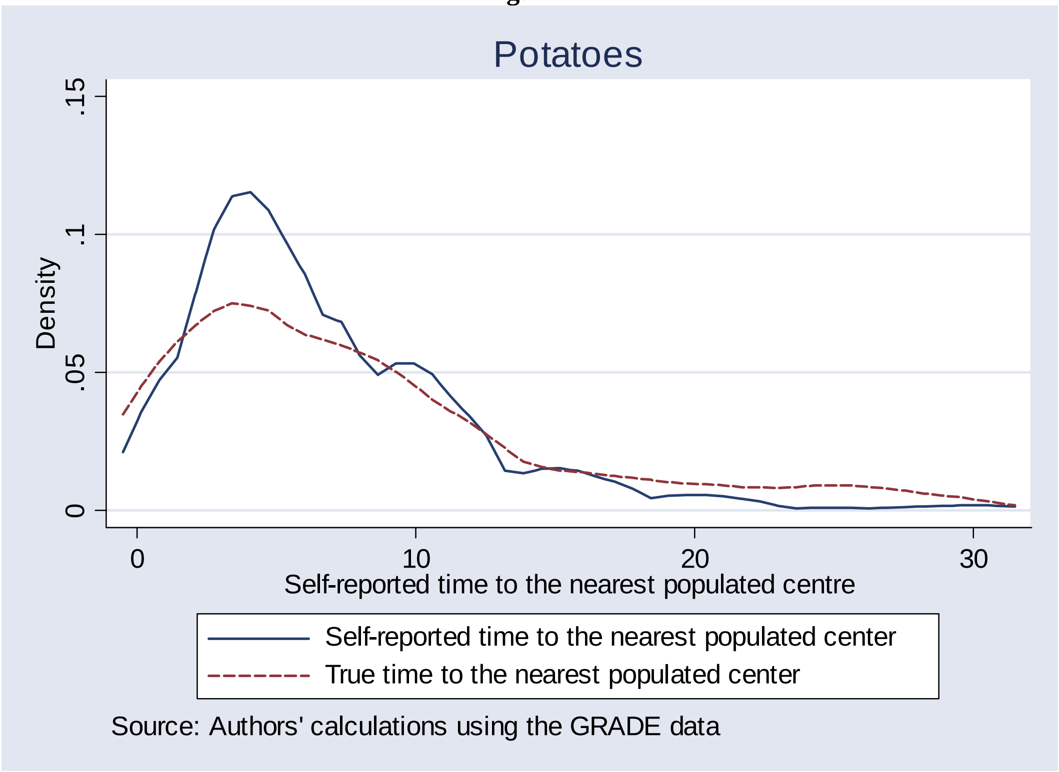

Figure 4

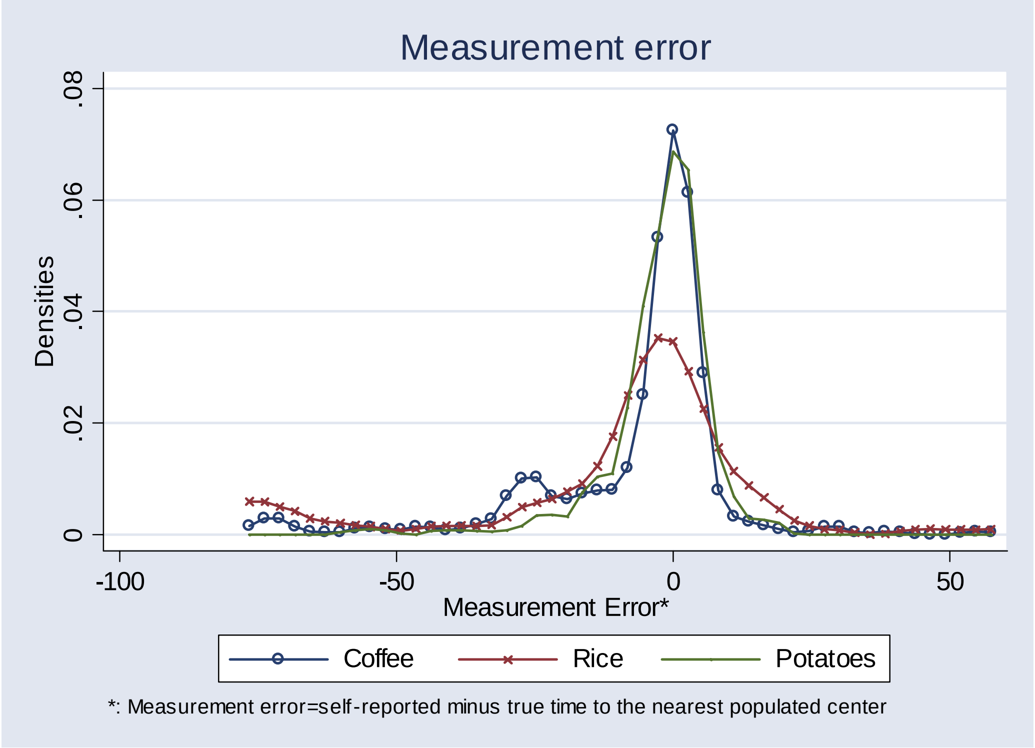

Figure A1

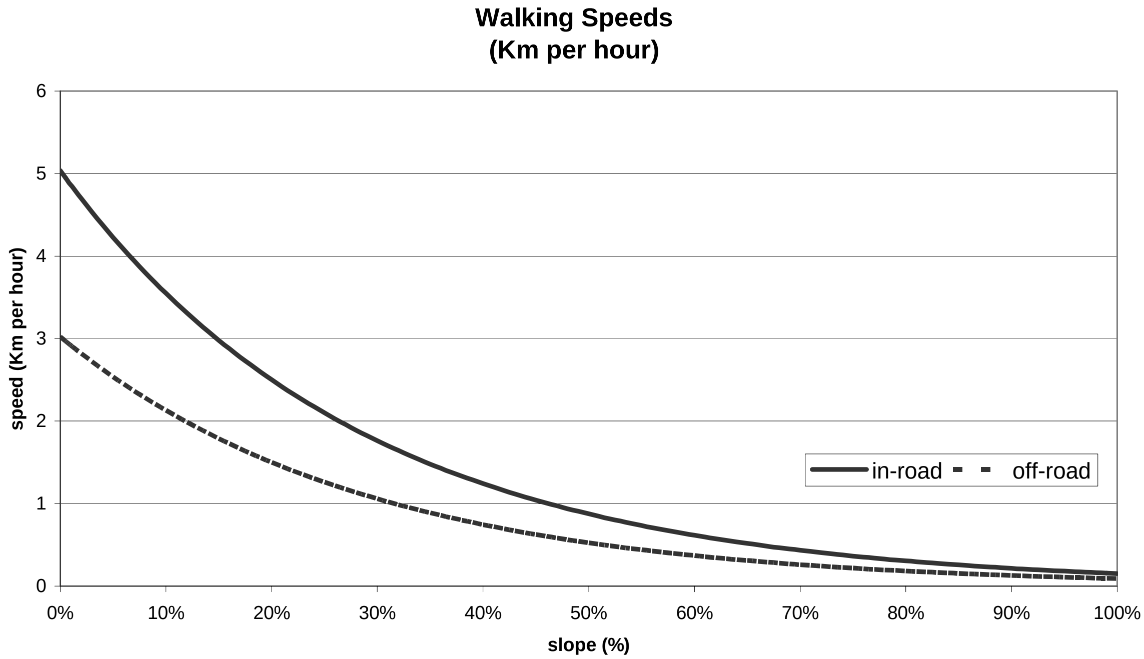

Figure A2 GIS Simulated Effective Trajectories for a Coffee Producer in Peruvian Amazon