WPS5550

Policy Research Working Paper 5550

Using Repeated Cross-Sections to Explore Movements in and out of Poverty

Peter Lanjouw
Jill Luoto
David McKenzie

The World Bank
Development Research Group
Poverty and Inequality Team
and Finance and Private Sector Development Team
January 2011

Policy Research Working Paper 5550

Abstract

Movements in and out of poverty are of core interest to both policymakers and economists. Yet the measurement of such movements has been limited in many countries due to the lack of panel data. In this paper, the authors consider a method whereby repeated cross-sections of household survey data can be analyzed in such a way as to allow inferences to be made about movements in and out of poverty. Their approach builds on the methodology used to construct poverty maps. They suggest that this method lends itself also to the development of pseudo-panels which can be used to study questions about poverty duration and mobility that are of interest to policy makers but that are rarely pursued empirically due to data constraints. They illustrate that the method at best offers insights into approximate *bounds* of mobility, but argue that these bounds can, under ideal circumstances, be narrow enough to yield useful insights. They test how well the method works by sampling repeated cross-sections from genuine panel data sets for Vietnam and Indonesia and comparing their method to the panel estimates. The results are sufficiently encouraging to offer the prospect of some limited, basic insights into mobility and poverty duration in settings where historically it was judged that the data necessary for such analysis were unavailable.

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

Using Repeated Cross-Sections to Explore Movements in and out of Poverty[*]

Peter Lanjouw, World Bank
Jill Luoto, RAND Corporation
David McKenzie, World Bank, BREAD and IZA

Keywords: Transitory and Chronic poverty; Pseudo-panels; Mobility.

JEL Codes: O15, I32.

[*] We are grateful to Chris Elbers, Martin Ravallion and Roy van der Weide, and seminar participants at Cornell, Minnesota, and the World Bank for useful comments. This paper represents the views of the authors only and should not be taken to reflect those of the World Bank or any affiliated organization.

"But the whole picture of poverty is not contained in a snapshot income-distribution decile graph. It says nothing about the vital concept of mobility: the potential for people to get out of a lower decile – and the speed at which they can do so."

Prime Minister David Cameron, October 2010[1]

1. Introduction

Recent decades have seen a mushrooming of evidence on poverty in developing countries. In the early 1990s, when the World Bank first attempted to estimate the extent of global poverty in its *World Development Report* on Poverty, 22 household surveys from 22 countries could be drawn on. Today, the World Bank's estimates use 700 surveys from 115 countries (Chen and Ravallion, 2010). The current evidence base allows not only for detailed, country-specific, estimates of poverty but, because for many countries there are several rounds of data corresponding to different time periods, also reveals how poverty has evolved over time. This evidence on the extent and evolution of poverty is central to assessments of development experience at the country level, and as input into formulation of policies and strategies for development.

While the available data provide rich opportunities for the analysis of poverty levels, lack of data and appropriate tools has meant much less attention has been devoted to answering a suite of important questions of central importance to both academics and policymakers relating to movements in and out of poverty. Evidence on poverty trends provides policy makers with important information about the aggregate direction of change over time. But the most commonly available data for the analysis of poverty – derived from cross-section surveys - provide only "snap-shots" of poverty. At best there may be multiple snap-shots available, corresponding to different moments in time, and permitting an assessment of aggregate poverty trends. Such data do not track individuals or households over time, and are therefore unable to reveal much about the duration of poverty experienced by specific households or population groups.

[1] Taken from a commentary "What you receive should depend on how you behave" in *The Independent*, October 10, 2010, http://www.independent.co.uk/opinion/commentators/david-cameron-what-you-receive-should-depend-on-how-you-behave-2102576.html

It is clear that an assessment of poverty will differ depending on whether poverty is experienced over long periods of time, or only fleetingly. Two countries with the same aggregate poverty rate may differ markedly in terms of the likely duration of poverty among the population. It may be that the same subset of the population is continuously poor in one country, while in the second many different people become briefly poor at different times. The consequences of such different experiences of poverty will differ, and the policy responses will also vary. For example, when there is chronic poverty among certain population groups, policy makers will likely seek to remove "poverty traps" that prevent the affected populations from escaping poverty. When poverty is generally transitory, but possibly widely experienced, the search may instead be for appropriate safety-nets that help to prevent individuals from falling, albeit briefly, into poverty and to provide a temporary cushion when they do.

Panel data are ideally suited to the analysis of income mobility – whether in and out of poverty, or across fractiles of the full distribution of income – as well as additional, related, subjects such as the study of vulnerability to poverty. Such panel surveys revisit the same individuals and/or households over time and thus permit comparisons of welfare across the same units of observation. But such surveys also pose challenges. Most notably, it can be complex and costly to field panel studies. In developing countries it may be particularly difficult to keep track of households over time, as they often move physically, and may also divide. Considerable effort and costs may be incurred in order to keep respondents in the sample. Those that do "drop-out" of the survey may affect representativity, because attrition may not be random. Because of these, and other, related, difficulties, panel surveys remain far less common than standard cross-section surveys, and as a result, the analysis of welfare dynamics also lags behind the kind of analysis that can be supported by cross-sectional data.

This paper introduces and explores a statistical methodology whereby two or more rounds of cross-section data are converted into a 'pseudo' panel that can support at least some analysis of welfare dynamics. The approach builds on an "out-of-sample" imputation methodology described in Elbers et al (2002, 2003) for small-area estimation of poverty (the development of "poverty maps"). A model of consumption (or income) is estimated in the first round of cross-section data, using a specification which includes only time-invariant covariates. Parameter estimates from this model are then applied to the same time-invariant regressors in the second survey round to predict an estimate of the (unobserved) first period's consumption or income for the individuals surveyed in that second round. Analysis of mobility can then be based on the actual consumption observed in the second round along with this estimate from the first round. Depending on how disturbances from the underlying imputation models are treated we designate one set of mobility estimates as an approximate upper-bound estimate; likely to be overstating the extent to which individuals or households in the population have entered and exited out of poverty. We claim that our second set of estimates produces an approximate lower-bound assessment of mobility.

The method is applied in two empirical settings: Vietnam and Indonesia. As we have genuine panel data in these applications, we are able to sample repeated cross-sections from the panel, construct our mobility estimates, and then compare them to those one would obtain using the actual panel data. In this way we are able to validate our method. We find that the "true" estimate of the extent of mobility (as revealed by the actual panel data) is generally sandwiched between our upper-bound and lower-bound assessments of mobility.[2] Our analysis reveals further that the gap between the upper- and lower-bound estimates of mobility is narrowed as the prediction models are more richly specified. We suggest that these results are sufficiently encouraging to warrant further exploration and experimentation with this methodology. We argue that the approach may be particularly promising if cross-section survey efforts were to incorporate into their design certain data collection elements that would strengthen the method's underlying prediction models.

2. Related Literature

The approach described in this paper is related to, but also distinct from, other proposals in the literature for developing pseudo-panels from multiple rounds of cross-section data.[3] Existing papers impose more structure to get point estimates of particular mobility measures, while we require fewer assumptions to achieve our bounds. Bourguignon, Goh and Kim (2004)

[2] Our bounds are "approximate" because sampling errors can in principle result in the "truth" lying outside the interval set by the bounds. Our bounds can thus be best understood in an asymptotic sense. We show below that in our data departures from strict bounding are very rare.

[3] A somewhat related literature considers how panel data on a subset of individuals can be used to infer chronic poverty for a larger sample (see Gibson, 2001).

develop a pseudo-panel based on 10 annual rounds of Wage Structure Survey (WSS) data from Korea. Their approach is based on techniques focusing on second-order moments in which individual earning dynamics can be analyzed with cross-sectional data under simplifying assumptions about individual earning dynamics. The method becomes more attractive as the number of cross-sections increases and is unlikely to work well, for example, with only two cross-sections. It also relies heavily on the functional forms assumed for earnings dynamics (an AR(1) process and log normality of earnings with no individual fixed effects) and will not deliver consistent estimates of mobility if these assumptions are violated.

Antman and Mckenzie (2007) employ a pseudo-panel approach as a means to sidestep well-known problems with non-classical measurement error in conventional panel analysis. They build pseudo-panels on the basis of cohort averages and tracking these cohorts through multiple rounds of cross-section survey data.[4] They can consistently identify one measure of (im)mobility – the projection of income or consumption on its lag – insofar as it arises from shocks to an AR(1) process, from demographic factors, or from cohort-level shocks. However, they miss mobility coming from transitory shocks to individuals. The bounds provided in this paper incorporate any movements into and out of poverty coming from such shocks, although we cannot distinguish whether the cause of such movements is transitory shocks or measurement error. Our aim here is merely to replicate what one would measure in a panel setting. We believe this will still be of use to policymakers interested in comparisons over time or across groups of individuals.

3. A New Approach to Measuring Movements into and out of Poverty with Repeated Cross-sections

For ease of exposition we consider the case of two rounds of cross-sectional surveys, denoted round 1 and round 2. The same approach used here can easily then be extended to examine dynamics in and out of poverty over additional rounds. We assume that both survey

[4] Most of the literature using pseudo-panels has followed this same approach of constructing cohort means and using these to form a panel at the cohort level (e.g. Deaton (1985), Moffitt (1993) and the review in Verbeek (2008)). As noted in the text, the use of cohort-means precludes the examination of income mobility within cohorts, which the approach proposed in our paper allows.

rounds are random samples of the underlying population of interest, and each consist of a sample of N households.

Let x_{i1} be a vector of characteristics of household i which are observed (for different households) in both the round 1 and round 2 surveys. This will include all time-invariant characteristics of the household such as language, religion, and ethnicity, and if the identity of the household head remains constant across rounds, will also include time-invariant or deterministic characteristics of the household head such as age, sex, education, place of birth, and parental education. Importantly it can also include time-varying characteristics of the household that can be easily recalled for round 1 in round 2. Thus variables such as whether or not the household head is employed in round 1, and his or her occupation, as well as their place of residence in round 1 could be included in x_{i1} if asked in round 2.

Then for the population as a whole, the linear projection of round 1 consumption or income, y_{i1}, onto x_{i1} is given by:

$$y_{i1} = \beta_1' x_{i1} + \varepsilon_{i1} \tag{1}$$

And similarly, letting x_{i2} denote the set of household characteristics in round 2 that are observed in both the round 1 and round 2 surveys, the linear projection of round 2 consumption or income, y_{i2} onto x_{i2} is given by:

$$y_{i2} = \beta_2' x_{i2} + \varepsilon_{i2} \tag{2}$$

Let z denote the poverty line. Then to estimate the degree of mobility in and out of poverty we are interested in knowing, for example, what fraction of the households which are above the poverty line in round 2 were below the poverty line in round 1. That is, we are interested in knowing quantities such as:

$$Pr(y_{i1} < z | y_{i2} > z) \tag{3}$$

3.1 Our Proposed Method

The prime difficulty facing us with repeated cross-sections is that we do not know y_{i1} and y_{i2} for the same households. We propose the following steps to overcome this problem and estimate quantities such as that in equation (3).

Step One: Using the sample of households surveyed in round 1, regress y_{i1}^1 on x_{i1}^1, where the superscript denotes that these are observations for households observed in round 1 only. From this OLS estimation obtain the OLS estimator of β_1', denoted $\hat{\beta}_1$, and also the OLS residuals:

$$\hat{\varepsilon}_{i1}^1 = y_{i1}^1 - \hat{\beta}_1{}' \, x_{i1}^1 \tag{4}$$

Step Two: For each household *i* observed in round 2, take a random draw with replacement from the empirical distribution of the residuals defined in equation (4) and denote it by $\tilde{\varepsilon}_{i1}^2$. Then use the OLS estimate $\hat{\beta}_1$ together with the known value x_{i1}^2 to estimate round 1 income or consumption as follows:

$$\hat{y}_{i1}^2 = \hat{\beta}_1{}' x_{i1}^2 + \tilde{\varepsilon}_{i1}^2 \tag{5}$$

Step Three: Calculate the movements into and out of poverty of interest, using $\hat{y}_{i1}^2$ in place of the unobserved y_{i1}^2. For example, estimate equation (3) by

$$Pr(\hat{y}_{i1}^2 < z | y_{i2}^2 > z) \tag{6}$$

Step Four: Repeat steps one through three R times, and take the average of (6) over the R replications. We use R=100 in our simulations below.

3.2 What Conditions Are Needed for Consistency?

The method described above will give us consistent estimates of the degree of movement into and out of poverty if the following two key conditions are satisfied:

Condition 1: The underlying population sampled is the same in round 1 and round 2. This condition ensures that as the sample size goes to infinity, $\hat{\beta}_1 \rightarrow_p \beta_1$.

Condition 1 will not be satisfied if the sampling methodology or measurement of consumption is changed from one round to the next, or if the underlying population changes through births, deaths, or migration out of sample. As in standard pseudo-panel analysis these assumptions will be best satisfied by restricting attention to households headed by people aged say 25 to 55. Analysis of mobility among households headed by those younger than 25 or older than 55 or 60 is more difficult since these ages are often when households are beginning to form, or starting to dissolve. Since income may be measured at the individual level, this is less of a concern for individual income mobility than for household consumption mobility.

Condition 2: ε_{i1} is independent of y_{i2}. Since we have defined ε_{i1} through a linear projection it is orthogonal to x_{i1} and thus to x_{i2}. This assumption therefore amounts to assuming that ε_{i1} is independent of ε_{i2}. If this assumption is satisfied, then the distribution of $\varepsilon_{i1}|y_{12} > z$ is the same as the unconditional distribution of ε_{i1}. This enables us to use the unconditional empirical distribution of the estimated residuals in step two.

Condition 2 will not be satisfied if ε_{i1} is correlated with ε_{i2}. There are two important types of cases where this could happen. The first is if the error term contains an individual fixed effect. In this case households which have consumption higher than we would predict based on their *x* variables in round 1 will also have consumption higher than we would predict based on their *x* variables in round 2. The presence of these fixed effects therefore tends to reduce mobility. Our method would then overstate movements into and out of poverty if this was the case. The second case is if shocks to consumption or income are non-transitory. We would expect most shocks to income that have some persistence to be positively correlated. For example, getting a new job or losing a job, or getting a pay rise. If consumption reacts to these income shocks, then consumption errors will also exhibit positive autocorrelation. This will again lead our method to overstate mobility. It seems less likely that the errors will be negatively correlated on average. For particular households we might see this reversion. For example, a household which lacks access to finance may have had to cut expenditure in round 1 in order to pay for a wedding in round 2. For such a household we would see them having lower consumption than their *x* variables would predict in round 1, and higher consumption than would be predicted for round 2. However, this is unlikely to occur for the majority of households at the same time, so we believe that in most cases the correlation of errors will not be negative. We will show this for the two panel data sets used in our analysis.

3.3 Upper and Lower Bounds on Mobility

On balance then we believe that condition 2 is likely to be violated in many data sets, and that the ε_{i1} will on average be positively correlated with ε_{i2}. As a result, our method is likely to *overstate* mobility, providing an *upper bound* on movements into and out of poverty.[5] A partial solution to this problem is to enrich the set of variables included in the vector *x*. It may be that we can span the household fixed effects by including detailed geographic variables or region fixed effects, along with a host of other time-invariant household characteristics. The inclusion of such characteristics will also mean we are controlling for shocks which occur for particular groups, such as members of the same occupation, or people from the same village. This should reduce the autocorrelation of the errors by removing group-level autocorrelation.

Secondly, we can attempt to provide a *lower bound* on mobility. Since our concern is overstating mobility due to failure to take account of the possible positive autocorrelation of errors, we can construct a lower bound by assuming instead that the prediction error for household *i* in round 1 is exactly equal to its prediction error in round 2, thereby assuming perfect positive autocorrelation. That is, for the group of households observed in round 2, we obtain a lower bound on mobility by the following steps.

Step One: Using the sample of households surveyed in round 1, regress y_{i1}^1 on x_{i1}^1, where the superscript denotes that these are observations for households observed in round 1 only. From this OLS estimation obtain the OLS estimator of β_1', denoted $\hat{\beta}_1$. Then for the sample of households surveyed in round 2, regress y_{i2}^2 on x_{i2}^2 and obtain the OLS residuals:

$$\hat{\varepsilon}_{i2}^2 = y_{i2}^2 - \hat{\beta}_2{}' x_{i2}^2 \qquad (7)$$

[5] A second reason why this method will overstate mobility arises from the fact that we are using the predicted value of β_1 instead of the true value. This prediction error means we are adding additional noise to our estimate of round 1 income or consumption, and this additional noise leads to an overstatement of mobility. However, as the sample size increases the estimate of β_1 converges to the true value, shrinking the variance of this prediction error to zero. Thus in large samples this second source of overstatement becomes less important.

Step Two: Use the OLS estimate $\hat{\beta}_1$ from round 1 together with the known value x_{i1}^2 to estimate round 1 income or consumption as follows:

$$\tilde{y}_{i1}^2 = \hat{\beta}_1{}' x_{i1}^2 + \hat{\varepsilon}_{i2}^2 \quad (8)$$

Step Three: Calculate the movements into and out of poverty of interest, using $\tilde{y}_{i1}^2$ in place of the unobserved y_{i1}^2. For example, estimate equation (3) by

$$Pr(\tilde{y}_{i1}^2 < z | y_{i2}^2 > z) \quad (9)$$

Note that we do not need to replicate these steps *R* times, unlike the procedure for obtaining the upper bound, since we are using household *i's* own prediction error, not drawing repeatedly from the empirical distribution of prediction errors.

Under the assumption that prediction errors are not negatively autocorrelated, these two approaches will then give an upper and lower bound to the extent of movements into and out of poverty.[6] Note that if we have a rich class of variables to condition on so that the autocorrelation in errors is close to zero, then the upper bound is our preferred estimate, since it will be consistent for the true level of mobility one would observe in a panel if conditions 1 and 2 hold. Secondly, note that the width of the bounds should shrink as the variance of the prediction errors falls. Thus using more *x* variables to better predict consumption levels should give narrower ranges for mobility.

3.4 A Note on Measurement Error

The methods developed here aim to estimate the same level of movements into and out of poverty that one would observe in the genuine panel. Of course some of the mobility in the genuine panel data is spurious, arising from measurement error. There are several approaches in the existing literature for ways to correct mobility measures for this measurement error (e.g. Glewwe, 2005; Antman and McKenzie, 2007; Fields et al. 2007). The basic idea underlying all of these approaches is to study mobility which can be related to the mobility of some underlying

[6] These are bounds in the sense that the true mobility should lie between them. However, in finite samples, the *estimate* of mobility from a genuine panel could conceivably lie outside these bounds due to sampling error. This is unlikely to be a concern for most applications.

variable – such as health, cohort characteristics, or assets. This is analogous to studying only the mobility which comes from the $\beta'x$ term and ignoring mobility which comes from ε. Such an approach could be pursued here also. However, it is not the purpose of our current exercise, which is to determine whether one can use repeated cross-sections to estimate the same level of mobility one sees in a panel, and whether the method is useful for showing which characteristics are associated with more movements into and out of poverty.

Note however that our estimates will still remain valid bounds for the true degree of mobility even under many types of measurement error. This is certainly the case for classical measurement error. Classical measurement error leads to an overstatement of movements in and out of poverty by adding a component of uncorrelated noise to the error term. Our upper bound assumes the entire error term is serially uncorrelated, so will still be an upper bound for true mobility in this case, while our lower bound, which assumes perfect positive autocorrelation, will still be a lower bound. Thus while our approach cannot solve the problem of obtaining accurate point estimates in the presence of measurement error, inferences based on the intervals generated by our bounds should still be robust to the presence of measurement error.

4. Data Sets and Variable Choice

To examine how well our method performs in practice we implement our procedure using genuine panel data from Vietnam and Indonesia. These two countries provide useful contrasts in which to examine movements into and out of poverty. Vietnam illustrates our methods during a period of rapid growth and large reduction in levels of poverty, during which movements out of poverty are more common than movements into poverty. Indonesia illustrates the method during a more static setting in terms of overall poverty levels, during which households are moving into and out of poverty at similar rates.

4.1 Data Sets

In Vietnam we use the Vietnam Living Standards Surveys (VLSS) of 1992/1993 and 1997/1998. These are both are very well respected data sources that cover a period of significant structural change and document an impressive reduction in poverty – from 58% in 1992/3 to

37% in 1997/8 at the national level. The VLSS 1992/1993 covers 4800 households and is representative at the national and regional levels. It was conducted between September 1992 and October 1993 and contains data on the schooling, health, employment, migration, housing, and fertility of household members, as well as household consumption and ownership of a variety of household durables. The second VLSS round was collected between December 1997 and December 1998, and contains essentially the same information as the earlier survey. Many components of the two surveys' questionnaires are identical. The 1997/1998 survey contains panel information on approximately 4300 of the original 4800 households interviewed in 1992/1993, but has a total sample size of 6002 households due to an expanded budget and sample design. It is also representative at the national and regional levels. We focus in this paper on the panel data set.

Our data for Indonesia come from the Indonesian Family Life Surveys that were fielded by the RAND Corporation as part of their Labor and Population Program in collaboration with UCLA and the University of Indonesia. We work in this study with the IFLS2 and IFLS3 rounds corresponding to respectively, 1997 and 2000. The IFLS2 interviewed 7,500 households and the IFLS3 survey interviewed 10,400. The IFLS surveys are remarkable in the extent to which efforts were made to follow households over time. The IFLS2 and IFLS3 managed to resurvey 94.4 and 95.3%, respectively, of the original 7224 households interviewed in 1993 for the IFLS1 round. As is the case for the VLSS, the IFLS surveys are multipurpose surveys that collect detailed information on a range of different topics – thereby permitting analysis of interrelated issues that are single-purpose surveys do not. Information on economic outcomes like income and labor market outcomes can be combined with information on health outcomes, education and a whole host of additional socioeconomic indictors. Finally, in 1997, the IFLS fielded, alongside the IFLS2 household survey, a community survey about respondents' communities and public and private facilities. The analysis below draws on both household and community level information.

We then split the VLSS and IFLS panels into two randomly drawn sub-samples (each representing half of the total sample). Call these sub-samples A and B respectively. Then we can use sub-sample A in the first round and sub-sample B in the second round as two repeated cross-sections which we then carry out our method on. We can then compare the mobility results obtained from using sub-sample A to impute round 1 values for sub-sample B to the results we would get using the genuine panel for sub-sample B.

For our basic analysis we use the national poverty line in Vietnam provided with the VLSS (corresponding to D1,789,710 in 1998 prices), and the Tornquist poverty line in the IFLS data set (corresponding to Rp 86,128.1 in 2000 prices)[7]. We then show later in the paper that our results are robust to the choice of poverty line used.

4.2 Variable Choice

Our approach is built on a linear projection of consumption in round 1 onto individual, household and community-level characteristics that are also present in the data for round 2. As described in Elbers, Lanjouw and Leite (2009) with respect to the Elbers, Lanjouw and Lanjouw (2002, 2003) poverty-mapping procedure, there is no theory to guide the specification of what is essentially a forecasting model. However, certain diagnostics can be looked to for guidance. In general one would want to look well beyond explanatory power (a higher R^2 would tend to reduce the variance of the prediction error) to consider also statistical significance of the parameter estimates $\hat{\beta}_1$ (in order to reduce model error and the resultant overstatement of mobility) and to pay attention as well to concerns about over fitting. In the literature on poverty mapping, regressors have typically been drawn from several broad classes of variables including demographic variables (household size, gender and age profiles of households, etc.); human capital variables; labor market variables (occupational profiles), access to basic services and infrastructure (electricity access, connection to a piped water network, etc.); housing quality variables; ownership of durables; and community and locality-level variables.

Central to the present application of this approach is the additional requirement that regressors in these models be time invariant. Some variables are obvious candidates, such as the ethnic, religious, or social-group membership of the household head. Other time-invariant variables can be readily constructed from the data, such as whether the household head was aged 15 or higher and educated at the primary school level by a particular moment in time. When

[7] We thank Kathleen Beegle and Kristin Himelein for help with the IFLS data.

retrospective data are collected the range of time-invariant variables can be greatly expanded. For example, whether or not I have a fridge in 1992 is a time-invariant variable, which I can use in prediction models if both the 1997 and 1992 surveys collect information on whether I had a fridge in 1992. Some retrospective variables, such as place of residence at the time of the last survey, are reasonably common in cross-sectional surveys. There are other variables, such as sector of work, education level, and occupation at the time of the past survey, that are not often asked retrospectively, but could easily be without too many concerns as to recall error. Context will also determine the choice of variables to use. If the main interest is on mobility in rural farming areas, one could presumably ask retrospective questions about land and major livestock holdings, and also condition on time-varying environmental variables like rainfall.

4.3 A Hierarchy of Prediction Models

We consider a hierarchy of prediction models which progressively employ more and more data that is sometimes, but not always, collected retrospectively. Since we have the actual panel data to work with, we can "force" regressors in round 2 to be time-invariant by using the round 1 values of selected variables. Clearly in a real-world application we would be dependent only on those variables collected during the second round, and would be concerned about possible recall error. We aim to select variables which we believe are likely to be recalled fairly accurately, and which could be asked retrospectively, with the aim of seeing what is feasible with our method.

Six classes of models are specified and estimated in each case, based on increasing numbers of explanatory variables. We start with a basic, "traditional" model that includes variables that are most obviously time-invariant and are typically collected in almost all standard surveys. We then experiment with richer specifications that rely more and more on a richer array of variables which might require more concerted efforts in cross-sectional data collection to collect retrospectively. The six models are built up progressively as follows:

1. (Basic Model) We begin with a sparse model, including only variables that can be readily judged as time-invariant. For example, we can include such regressors as the gender of the head, age of the household head (defined in round 1 year), birthplace of the head

(rural/urban), whether the head ever attended primary school, the education level of the head's parents, and the head's religion and ethnicity.

2. We then introduce locational dummies such as urban/rural, or regional, dummies to measure where the household was living at the time of the first round survey. Most multipurpose surveys with a migration module would collect the information needed to allow these variables to be constructed, and even without a specific migration module, it is common to ask where households were living five years ago, which would work for the timing of the Vietnam surveys. Smith and Thomas (2003) examine how successfully Malaysian households can recall migration histories, with their results suggesting recall can be accurate, particularly for moves which are not very local or very short in duration.
3. Next, "community" variables are added, defined at the village level. In the case of Indonesia, these come from the community-level survey from 1997 and are inserted into both the IFLS2 and IFLS3 household surveys. For Vietnam, these variables are from the census and are commune-level means of a host of descriptors but are time-invariant by construction. Once the retrospective location is identified (as per model 2), the use of such variables depends only on the availability of such auxiliary data, and not on further recall per se.
4. We then add variables describing a household head's sector of work as well as richer measures of the education of the household head. At this point we clearly start to lean more heavily on our ability to explicitly insert round 1 values of these variables into the round 2 data. However, information on these variables could probably be easily collected on a retrospective basis. Indeed it is not uncommon for surveys to ask when educational qualifications were achieved, and retrospective work histories have been collected in a number of labor surveys.
5. Further demographic variables that we force to be time-invariant are then added - such as household size and the number of children aged under 5. These would possibly be more difficult to collect retrospectively if household composition is very fluid, especially if the time interval between survey rounds increased. Nonetheless, it is not uncommon for surveys with a migration focus to ask about all individuals who have lived in the household in the past five years, and our impression is that households in many societies are able to recall such information relatively accurately.

6. (Full model) Finally, we include a number of variables describing a household's assets and housing quality at the time of round 1 - such as ownership of specific consumer durables like a TV and motorcycle, and the type of roofing and flooring material the household had. Including these variables increases the predictive power of the consumption models significantly. Such variables are not commonly collected in retrospective fashion in large multipurpose surveys, but they have been collected in some specific survey contexts. For example, de Mel, McKenzie and Woodruff (2009) ask Sri Lankan business owners and wage workers questions on whether their family owned a bicycle, radio, telephone, or vehicle when they were aged 12, and on the floor type their household had then. Individuals were able to recall such information relatively easily, although further work is needed to test how accurate such recall is. Berney and Blane (1997) offer some encouraging findings from a small sample in the U.K., showing high accuracy recall of toilet facilities, water facilities, and number of children in the household over a 50-year recall period.

We estimate these models for log consumption per capita. We only use levels of the variables indicated above, but one could additionally enrich the models by including interactions (e.g. allowing the predictive impact of education for consumption to vary with region, sex of household head, etc.). The precise models used for the upper and lower bound estimates for model 1 (the "basic model") and model 6 (the "full model") are presented in Appendix 1.The basic model has an R^2 of 0.20 in Indonesia and 0.13 in Vietnam, while the full model has an R^2 of 0.42 in Indonesia and 0.53 in Vietnam. The greatest increments in explanatory power, for both Vietnam and Indonesia, arise with the move from the "basic model" to one including locational dummies, and then again with the addition of variables describing a household's asset position and housing quality. It is clear that addition of a progressively richer set of controls makes a large difference to the share of variation in log consumption explained by the model.

5. Results

5.1 How Well Do We Do at Predicting Consumption and Poverty in the Cross-section?

We first examine how well the model specifications employed here allow us to predict levels of poverty. As described above, we estimate the consumption models with one half of the data for round 1 (1992 in the case of Vietnam, and 1997 in the case of Indonesia). We then predict consumption and poverty in the other half of the data for round 1 and compare predicted poverty against directly measured poverty for the second half of the round 1 data.

Table 1 summarizes the results of fitting the basic and full models in Vietnam and Indonesia. Because directly measured poverty (called "Truth") is estimated with sampling error (and predicted poverty is associated with both sampling and prediction error) we cannot expect estimates to coincide perfectly. Rather, we are just interested to see whether predictions appear reasonably close to the "Truth". Table 1 reveals that the predicted headcount rates based on our various models are generally within the broad range of where we would expect them to be. Predicted poverty based on the "upper bound" method and the "full" specification lies within the 95% confidence interval of the "Truth" in both Indonesia and Vietnam.[8] This accords well with experience of applying the Elbers et al. (2002, 2003) method for small-area estimation purposes. In those applications the methodology pursued most closely resembles the "upper bound", "full", approach here, and it is generally found that predicted poverty rates (calculated in the population census) closely track survey estimates at the broad-stratum level (see Elbers et al. 2002, Demombynes, Elbers, Lanjouw, Lanjouw, Mistiaen and Ozler, 2004). The basic model has less predictive power, leading to wider intervals.

Table 1: Poverty Headcount

		Lower Bound		Truth		Upper Bound	
Data Source:	Round 1:	Basic	Full	95% CI		Basic	Full
IFLS	1997 Poverty Rate (P0):	0.149	0.159	0.145	0.188	0.164	0.150
VLSS	1992 Poverty Rate (P0):	0.611	0.592	0.597	0.682	0.558	0.578

[8] Note that the phrases "upper bound" and "lower bound" pertain to their bounds on mobility, not to their bounds on levels of poverty.

Following the discussion in section 2.2, we are interested to check whether round 1 errors and round 2 residuals from projecting consumption onto the *x* variables are correlated. Our claim that our "upper-bound" approach gives an upper bound is predicated on the assumption that errors are either uncorrelated or positively correlated. Table 2 shows that, indeed, as we move from class 1 to class 6, the correlation of residuals is positive in all cases but that it declines as the model specifications become richer. This is consistent with the inclusion of a richer set of regressors being able to capture some correlated shocks and span some of the individual fixed effects. Note that we would typically expect the autocorrelation to be smaller the longer is the interval length between surveys. However, in our context, the correlations can still be reasonably large, even with three to six years between surveys. This suggests in most applications the errors will be positively autocorrelated.

Table 2: Correlation between Round 1 and Round 2 Residuals

	1	2	3	4	5	6
Indonesia	0.474	0.466	0.464	0.452	0.408	0.348
Vietnam	0.653	0.575	0.563	0.539	0.523	0.420

Columns 1-6 build increasingly rich models of consumption.

5.2 How wWell Does This Approach Do at Estimating the Overall Rate of Movements into and out of Poverty?

We turn, now, to one of the central questions in our study, namely whether analysis of duration of poverty, and mobility in and out of poverty, based on our 'pseudo-panel' data, can deliver results approximating the findings one would obtain with genuine panel data. Table 3 presents our results. In general, as we expected, the lower bound estimates tend to underestimate mobility (overstating the extent to which people remain poor or remain non-poor and understating movements into and out of poverty) and the upper bound estimates overestimate mobility. The "truth" tends to lie about midway between these bounds. We find thus that our two approaches do indeed present bounds within which the "truth" can be observed.

What is particularly encouraging is that these bounds are not terribly wide. For example, using the full model, our bounds would suggest that between 3 and 11 percent of households in Indonesia, and between 27 and 34 percent of households in Vietnam moved out of poverty between the two rounds. Analysis based on the genuine panel data suggest that the truth lies somewhere in the 7-9% range in Indonesia, and 26-32% range in Vietnam. It thus seems that if only two cross-sections had been available in both of these countries, an assessment of entry and exit from poverty based on the method outlined here would have provided a fairly similar assessment of mobility as that which derives from proper panel data

Table 3: Poverty Dynamics from "Pseudo" Panel and Actual Panel Data

Indonesia: Full Set of Results from Class 1 ("Basic") to Class 6 ("Full") Models

	Lower Bound Estimates							**Upper Bound Estimates**					
1997, 2000 Statuses	1	2	3	4	5	6	**Truth**	1	2	3	4	5	6
Poor, Poor	0.115	0.112	0.103	0.102	0.106	0.106	0.058	0.031	0.030	0.027	0.027	0.032	0.035
							(0.005)						
Poor, Nonpoor	0.016	0.021	0.015	0.017	0.031	0.030	0.076	0.133	0.135	0.123	0.122	0.120	0.115
							(0.005)						
Nonpoor, Poor	0.020	0.024	0.032	0.034	0.030	0.030	0.076	0.105	0.105	0.108	0.108	0.103	0.100
							(0.005)						
Nonpoor, Nonpoor	0.847	0.843	0.848	0.847	0.832	0.833	0.781	0.730	0.729	0.741	0.741	0.743	0.748
							(0.008)						
R-squareds	0.197	0.23	0.23	0.243	0.338	0.42							

Note: R-squareds are calculated for opposite halves of the total 1997 sample.

Vietnam: Full Set of Results from Class 1 ("Basic") to Class 6 ("Full") Models

	Lower Bound Estimates							**Upper Bound Estimates**					
1992, 1998 Statuses	1	2	3	4	5	6	**Truth**	1	2	3	4	5	6
Poor, Poor	0.360	0.360	0.360	0.360	0.359	0.322	0.312	0.221	0.246	0.248	0.254	0.258	0.237
							(0.008)						
Poor, Nonpoor	0.241	0.250	0.236	0.238	0.237	0.274	0.292	0.337	0.317	0.311	0.307	0.301	0.341
							(0.008)						
Nonpoor, Poor	0.000	0.000	0.000	0.001	0.001	0.039	0.043	0.139	0.114	0.113	0.106	0.102	0.123
							(0.004)						
Nonpoor, Nonpoor	0.398	0.389	0.403	0.402	0.402	0.366	0.354	0.303	0.322	0.328	0.333	0.339	0.299
							(0.009)						
R-squareds	0.133	0.32	0.35	0.38	0.433	0.53							

Note: R-squareds are calculated for opposite halves of the total 1992 sample.

The results also illustrate the importance of being able to fit more detailed models to predict consumption, with narrower bounds for the full model than the basic model. Table 3 shows how the bounds generally narrow as more variables are added to the model. For example, the bounds for the proportion of the population falling into poverty in Indonesia between 1997 and 2000 are (0.020-0.105) using the basic model, (0.024-0.105) using model 2, (0.032-0.108) using model 3, (0.034-0.108) using model 4, (0.030-0.103) using model 5, and (0.030-0.100) using the full model. In both countries it is the inclusion of locational variables to get to model 2, retrospective demographic variables to get to model 5, and especially the inclusion of the retrospective household asset variables to get to the full model that most increase the share of variation explained by the regressors and the greatest reduction in the size of the bounds. Efforts to collect retrospective data so as to be able to enrich the model specification thus do appear to be important.

5.3 How Well Does This Approach Do at Describing the Profile of Mobility in and out of Poverty?

Interest in the analysis of mobility extends beyond an assessment at the aggregate level of the percentage of the population that has exited poverty, entered poverty, or retained welfare status over time. For policy makers the identification of population sub-groups that have seen significant upward or downward mobility, or that are relatively less mobile, is likely to be of particular interest. Figures 1-4 illustrate in turn for Vietnam and Indonesia how different population sub-groups have evolved over time based on both true panel data analysis and on analysis of the 'pseudo-panels' constructed on the basis of the upper-bound methodology using the full model specification.

Figure 1 considers exit from poverty between 1992 and 1998 in Vietnam, for population groups defined in terms of region of residence, ethnicity, urban or rural sector, and broad occupation status (household head engaged in agricultural sector or non-agricultural sector). We plot the likelihood of exiting poverty estimated by our upper bound approach against the actual occurrence of such movements in the genuine panel data. This enables one to see visually how similar the profiles of mobility are using the repeated cross-sections to what would be obtained using a genuine panel.

Observe first that the population subgroups follow the line of equality fairly closely indicating that an assessment of the likelihood of exiting poverty based on an analysis of pseudo-panel data is not markedly different from the assessed likelihood that would derive from analysis of actual panel data. Second, because we are plotting the upper bound method which overstates mobility, we would expect pseudo-panel estimates of the rate of poverty exit to lie above the line of equality. In Figure 1 this is broadly the case, although not universally so: the estimated exit of poverty amongst the urban population, and the populations of regions 4 and 6, based on pseudo-panel data, is slightly below the estimated based on true panel data. As has been noted above, all estimates of mobility are subject to sampling and/or prediction error and so some deviation from the expected pattern is not entirely surprising. In 10 out of 13 cases, the upper bound pseudo-panel analysis does indeed result in a somewhat higher assessment of poverty exit than true panel data would suggest.

Third, we can see from Figure 1 that with relatively few exceptions, comparisons of the rate of poverty exit between different population groups would be robust across our two types of data. Figure 1 illustrates, for example, that the percentage of the urban population who exited from poverty is considerably lower than the corresponding percentage of the rural population (not surprising given the lower overall poverty rates in urban areas), and this conclusion holds, irrespective of whether this analysis is based on pseudo-panel or genuine panel data. Similarly the likelihood of poverty exit for the population in region 2 is also higher than the population in region 4 or region 1, and the likelihood of poverty exit amongst the Kinh households is also higher than non-Kinh households, irrespective of methods. But agreement across methods is not universal: the likelihood of poverty exit for the population residing in region 7 is judged to be higher than in region 1, in the case of pseudo-panel analysis, but the reverse would be judged to be the case when analyzing actual panel data. Nevertheless, this is a case where the confidence intervals for the two regions overlap in genuine panel data, so our precision in distinguishing differences in poverty exit rates between them is not high even with the actual panel data.

Figure 2 turns to an analysis of entry into poverty in Vietnam, again comparing the upper bound approach with the full model to the genuine panel estimates. Note that few households entered into poverty in Vietnam during this period of rapid growth, so the scale is different from Figure 1. Here, the upper bound estimates of entry unambiguously overstate mobility. However, again, estimates of entry into poverty line up fairly closely along the 45-degree line, indicating broad agreement between the two methods. And again, assessments of the relative mobility of different population groups would be fairly similar across the two methodologies.

Figure 1

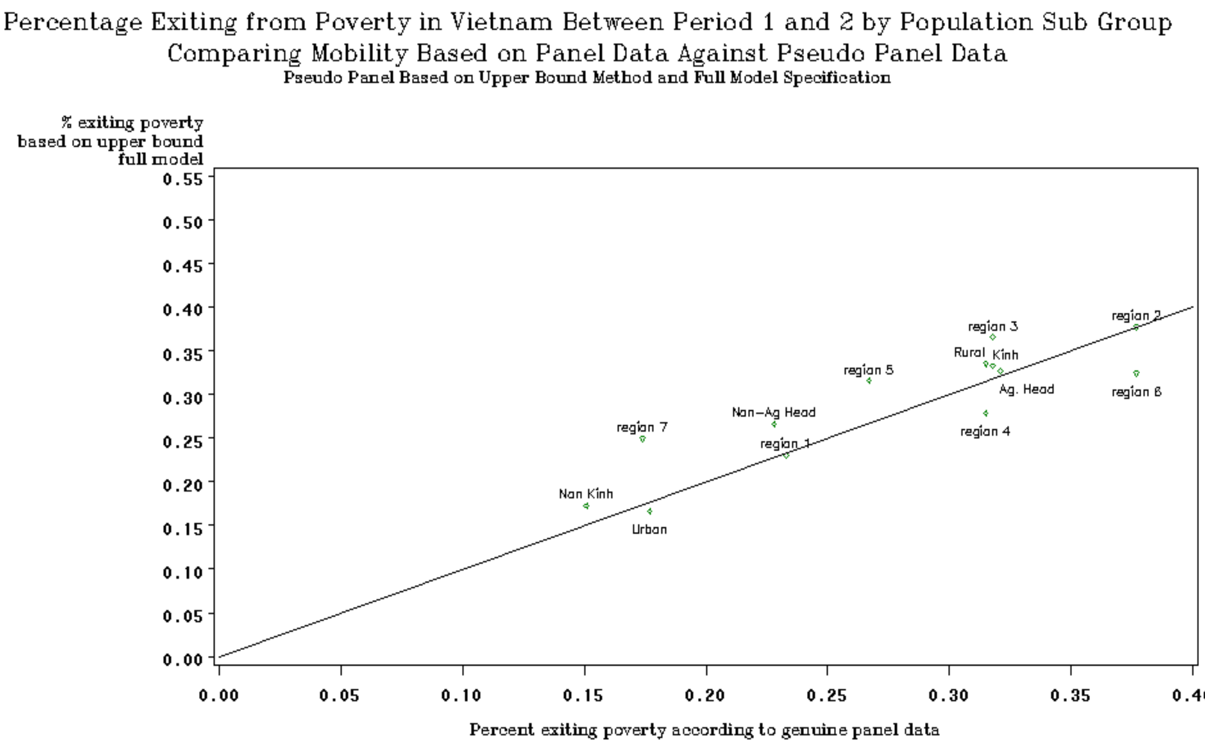
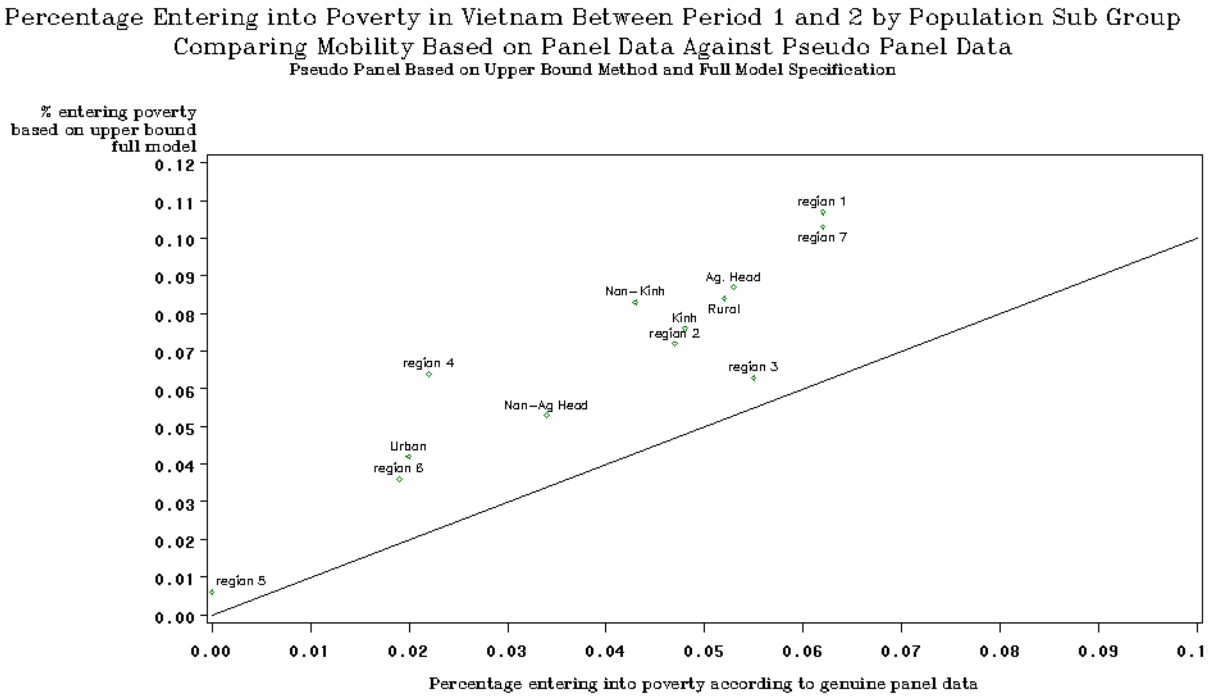


Figure 2



Figures 3 and 4 present the same analysis for Indonesia. In this case there are 21 possible population groups under consideration (compared to 13 groups in the case of Vietnam). The patterns are broadly similar to what was observed in the case of Vietnam: the pseudo-panel data (based on the upper bound method with the full model) overstate mobility, but to a moderate degree, and assessments of poverty exit, or entry, would not differ markedly between pseudo-panel based analysis or analysis of true panel data. Again, the broad relative "profile" of mobility in- and out- of poverty across population groups would be reasonably robust across methods.

Tables 4 and 5 summarize the possible comparisons between panel-based analysis and pseudo-panel analysis. We correlate estimated likelihoods of poverty exit and poverty entry across different population groups based on the two alternative approaches – considering, here, four variants of pseudo-panel based analysis (lower bound versus upper bound, basic versus full-model specifications). Pearson correlation and spearman rank-correlations are calculated. As expected, the pseudo-panel analysis comes closest to the "truth" in both Vietnam and Indonesia, and for both the case of poverty exit and poverty entry, when the upper bound methodology is applied in combination with the full model specification. As noted earlier, this approach is unlikely to be available in the absence of explicit efforts during collection of data in the second period to collect retrospective information on household demographics, occupations and ownership of assets in the past. In the absence of such retrospective information, analysis would likely have to be based on more basic model specification, resulting in weaker explanatory power and correspondingly more tentative estimates of mobility. As revealed in Tables 4 and 5, insights into sub-group mobility patterns derived from such a basic model would not be terribly robust.

Figure 3

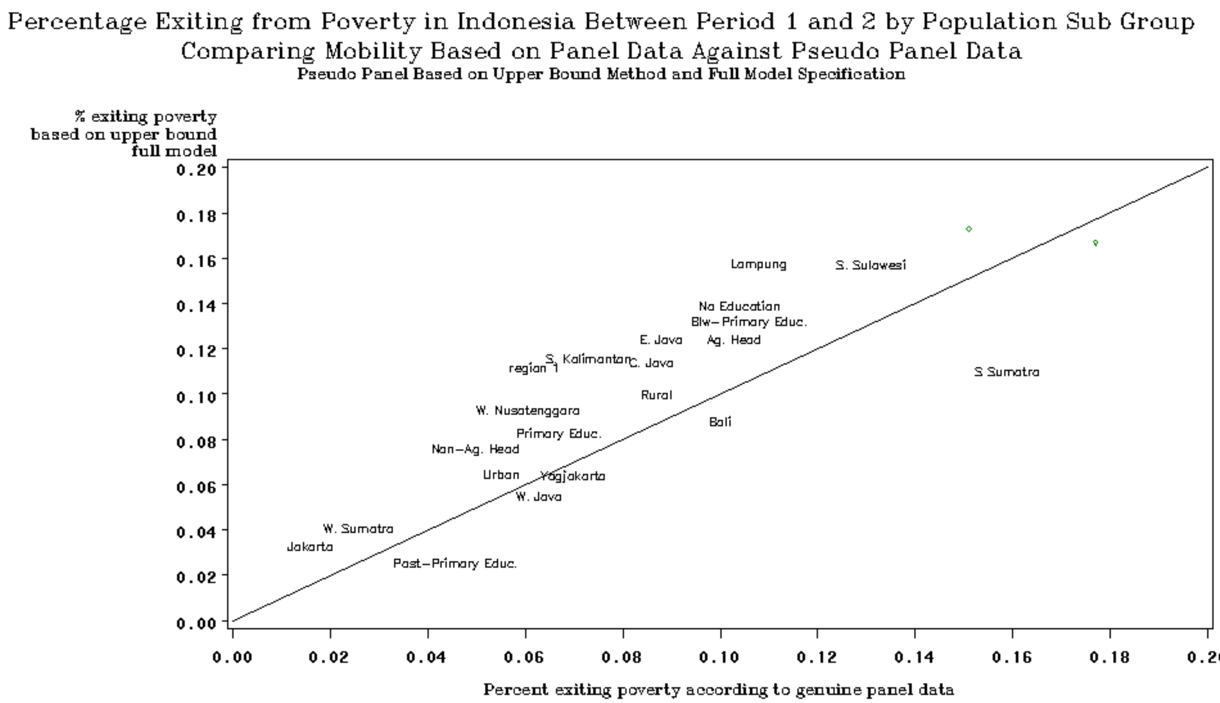


Figure 4

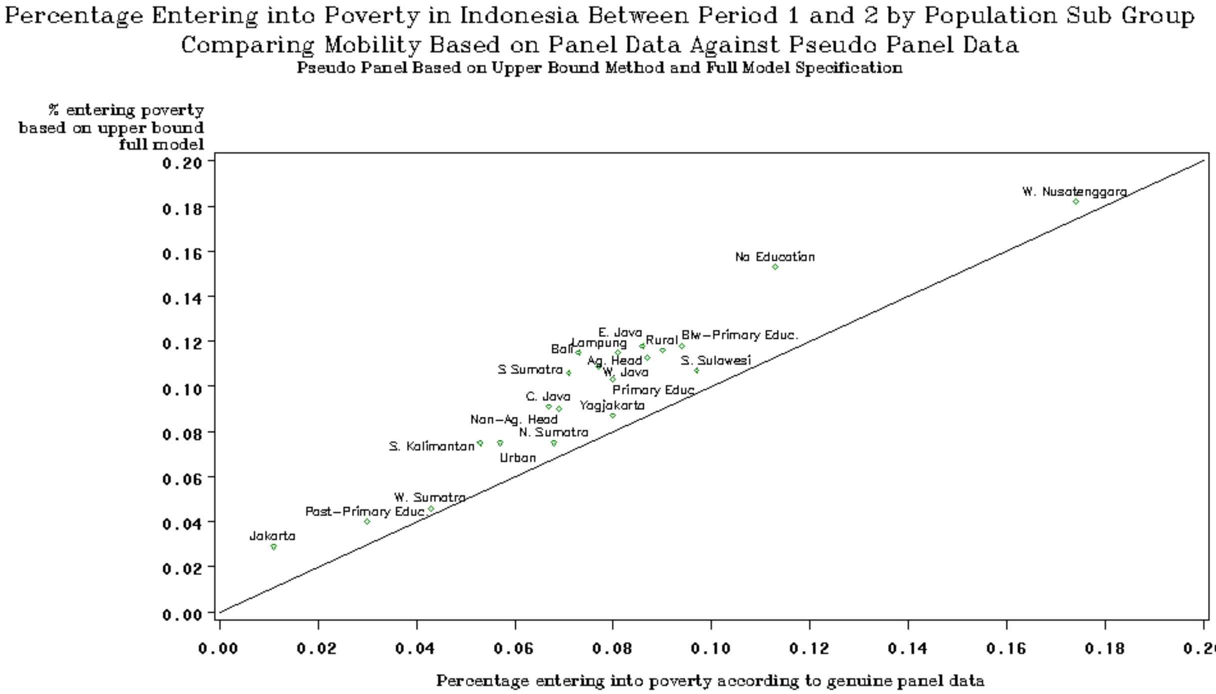


Table 4

Correlation of Panel-Based Probability versus Pseudo-Panel Based Probability of Exiting Poverty Across Different Population Sub-groups

	Vietnam (13 Population Sub-Groups)		Indonesia (21 Population Sub-Groups)	
	Pearson Correlation	Spearman Rank Correlation	Pearson Correlation	Spearman Rank Correlation
Lower bound method, basic model	0.399	0.480	0.331	0.388
Lower bound method, full model	0.650	0.770	0.189	0.170
Upper bound method, basic model	0.271	0.050	0.491	0.557
Upper bound method, full model	0.886	0.833	0.787	0.824

Table 5

Correlation of Panel-Based Probability versus Pseudo-Panel Based Probability of Entering Poverty Across Different Population Sub-groups

	Vietnam (13 Population Sub-Groups)		Indonesia (21 Population Sub-Groups)	
	Pearson Correlation	Spearman Rank Correlation	Pearson Correlation	Spearman Rank Correlation
Lower bound method, basic model	0.00	0.00	0.811	0.581
Lower bound method, full model	0.025	0.056	0.715	0.636
Upper bound method, basic model	0.862	0.837	0.922	0.880
Upper bound method, full model	0.920	0.856	0.945	0.898

5.4 Robustness to Choice of Poverty Line

The preceding analysis has all been based on one particular poverty line. The question arises as to whether the approach described here is also successful in bounding true mobility when alternative poverty lines are considered. A related question is whether the tightness with which our bounds "sandwich" the truth is constant for different values of the poverty line. We investigate these questions by calculating upper and lower bounds on mobility, as well as the truth, for the set of poverty lines spanning the range of possible base year poverty rates from 0 to 100 percent. Figures 5-8 illustrate our results in terms of the fraction of the population who escape poverty and the fraction who remain in poverty, in turn for Vietnam and Indonesia.

The VLSS "true" panel data indicate that the share of the population able to escape poverty is low when the base year poverty line (and hence aggregate poverty) are sufficiently low (Figure 5). As the poverty line increases in value, a larger share of the base year population is considered poor and the percent that escapes poverty also rises. As the poverty line continues to rise an increasing fraction of the base year population is counted as poor and eventually the share of that underlying population that manages to escape poverty starts to decline. When the line is sufficiently high the whole population is poor and remains poor. Figure 5 shows that the inverted U-curve pattern traced out by the VLSS panel data is tracked fairly closely by our lower and upper bound pseudo-panel estimates of mobility out of poverty. Allowing for some overlap and crossing attributable to statistical uncertainty, the bounds do "sandwich" the truth over the full range of possible poverty lines. In fact, in the Vietnam case, the lower bound estimates appear to track the truth more closely than the upper bound estimates. Figure 5 also illustrates that the gap between the upper and lower bound estimates is at its widest when around half of the base-year population is considered poor, and also the largest share of the population is able to escape poverty. At more extreme poverty lines, the bounds are much closer together, pointing also to much lower rates of mobility out of poverty.

Figure 6 present the analogous picture of mobility out of poverty for Indonesia. Again, an inverted U curve is traced out, and again the "truth", as revealed from actual panel data, lies between the upper and lower bound estimates deriving from our approach. In Indonesia, rates of poverty exit are much lower than in Vietnam across all possible poverty lines (consistent with an overall picture of less economic dynamism and aggregate poverty reduction in Indonesia during the period we are considering).

Figure 5: Estimates of Mobility Out of Poverty for Alternative Poverty Lines: Vietnam

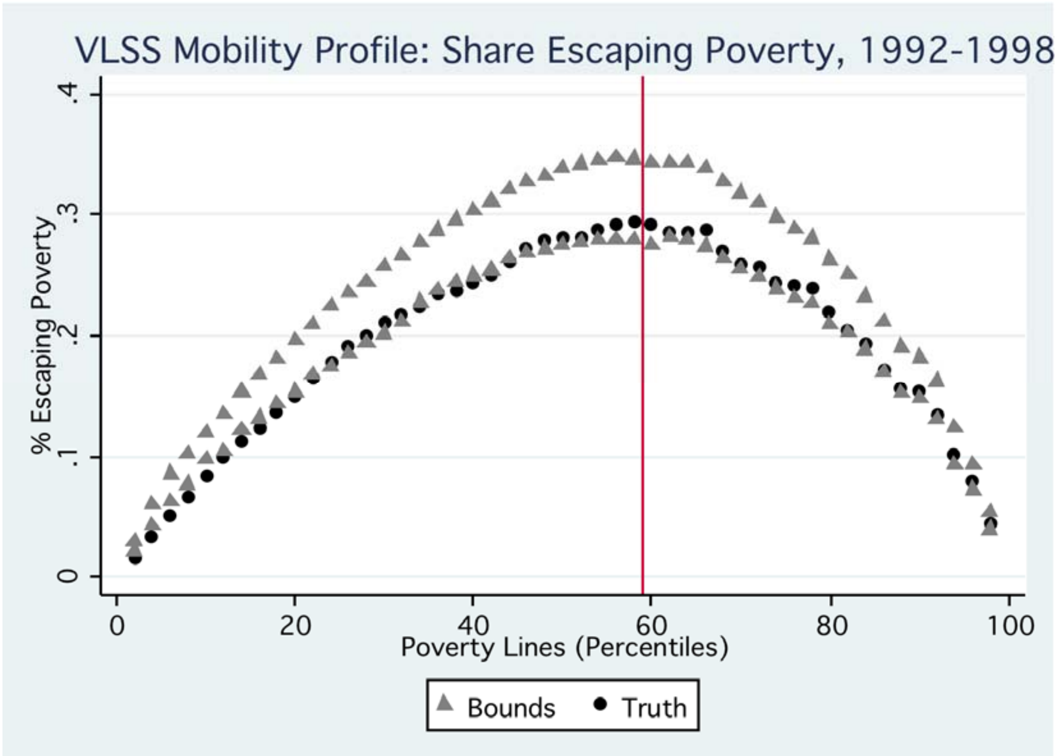
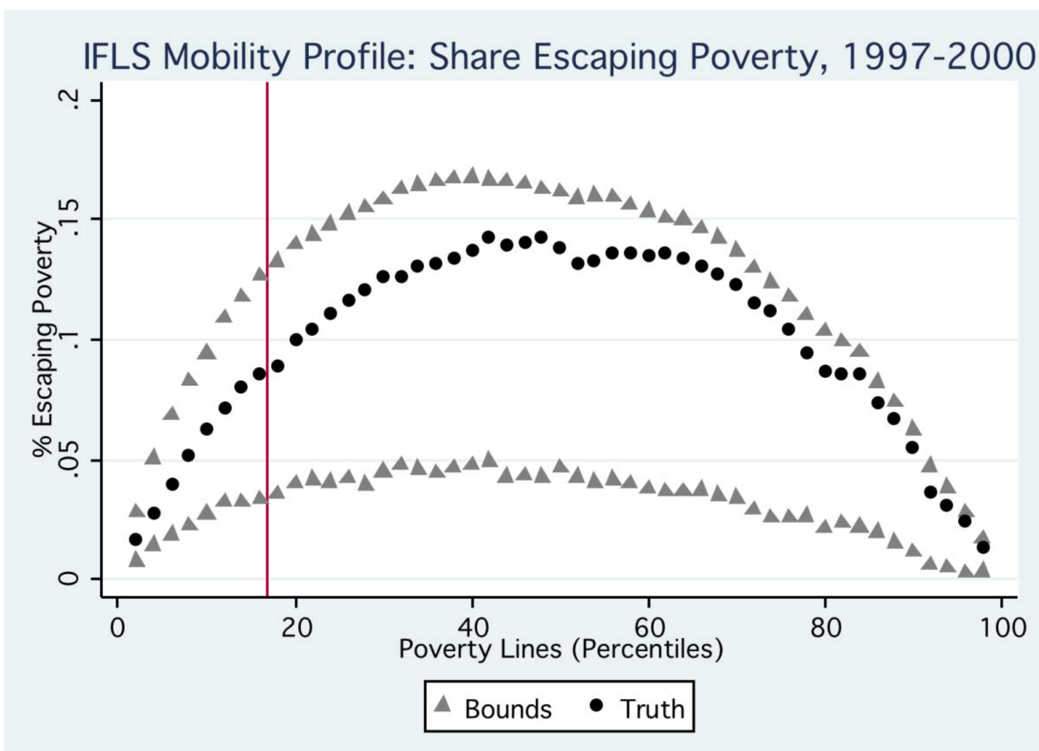


Figure 6: Estimates of Mobility Out of Poverty for Alternative Poverty Lines: Indonesia



In Figures 7 and 8 we consider the counterpart case of immobility out of poverty. In both Vietnam and Indonesia, as the poverty line increases, correspondingly larger shares of the underlying population are counted as chronically poor. Again, in both countries the pseudo-panel estimates produced here bound the "truth" quite well across the full range of possible poverty lines with, again, the largest gaps between the upper and lower bound estimates occurring around the mid-range of possible poverty lines (when around half the base-year population is counted as poor).

In sum, our approach is found to work well for the full possible range of poverty lines that might be specified, and we find that our bounds are, indeed, upper and lower bounds to the "truth" irrespective of where the poverty line is drawn.

Figure 7: Estimates of Prolonged Poverty Duration for Alternative Poverty Lines: Vietnam

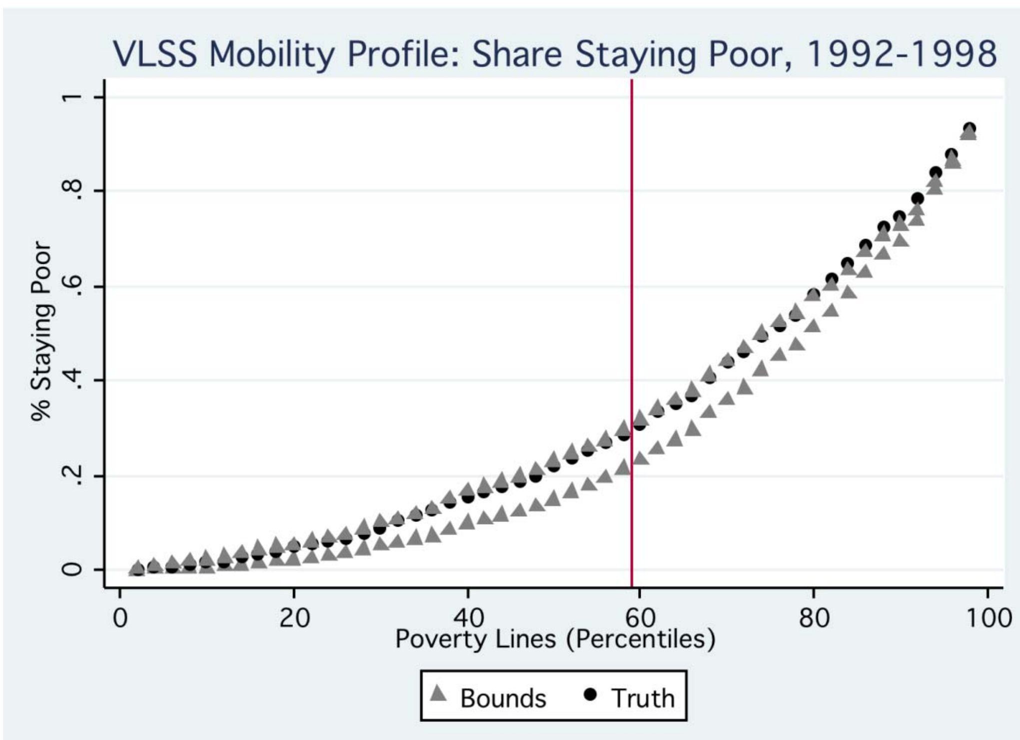
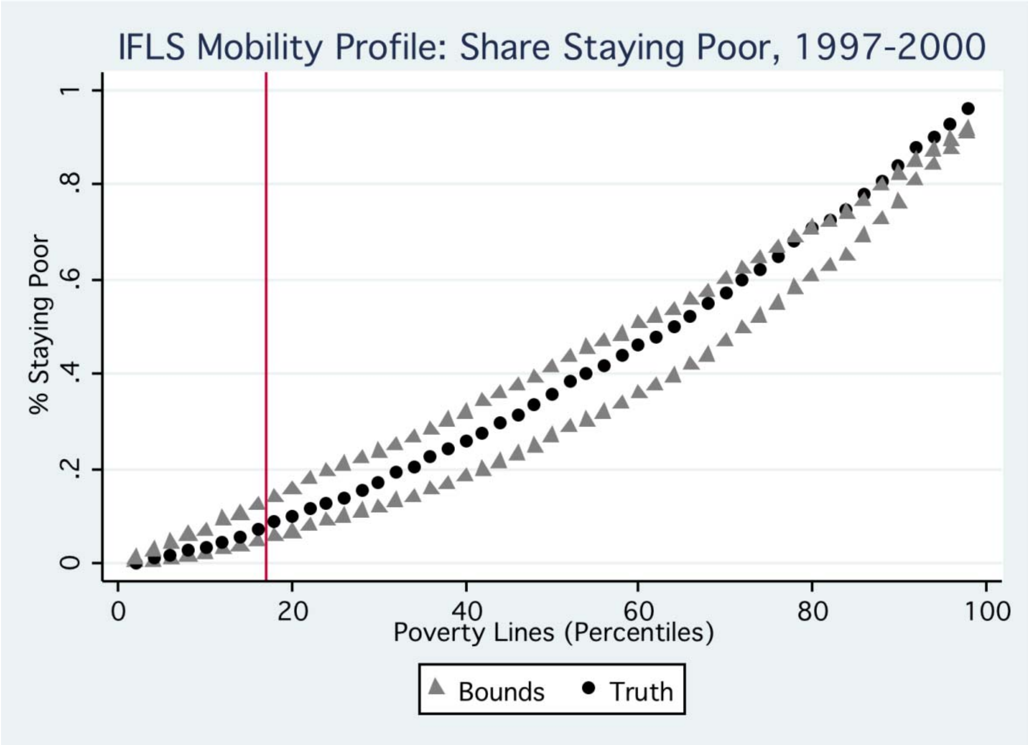


Figure 8: Estimates of Prolonged Poverty Duration for Alternative Poverty Lines: Indonesia



5.5 Extensions for Future Research

The approach presented here is intended to demonstrate how reasonable bounds for mobility can be obtained under fairly unrestrictive assumptions. As is the case with bounding approaches in general, narrower bounds can be obtained by imposing more structure. We suggest two avenues for doing this which researchers may wish to explore further in future applications. First, our upper bound is obtained by randomly sampling an ε_1 from the empirical distribution over the entire sample. If there is heteroskedascity in the ε_1, then one could improve on this by expanding on the approaches used in cross-sectional work, either by sub-sampling the ε_1 from within strata (e.g. young rural households) as in McKenzie (2005) or from explicitly modeling the heteroskedascity as in Elbers et al. (2003). Second, our bounds are based on the extreme cases of error autocorrelation of zero or one. If supplementary information is available, or one wishes to impose enough distributional assumptions to estimate this autocorrelation, it seems likely that the bounds could be narrowed by incorporating this information.

6. Conclusions and Future Directions

Genuine panel data are still rare in the developing world, and when they are available, it is often for a relatively small sample, with limited or infrequent duration, and in some cases, occurs with significant attrition. This has limited the feasibility of constructing even the simplest descriptions of movements in and out of poverty for most countries. Yet policymakers and researchers do care about such movements (as evidenced by David Cameron's quote in the introduction), and most countries do contain somewhat regular repeated cross-sectional surveys of income or consumption. In this paper we have developed a method for using this existing data to provide some bounds on the extent of movements into and out of poverty, and results from both Indonesia and Vietnam suggest these bounds can be made narrow enough in practice to make the estimates useful. Preliminary evidence to support this can be seen by new efforts underway to use the methodology developed in this paper to systematically examine poverty dynamics in a number of Latin American countries.[9]

However, the success of our approach depends on how well we can predict consumption (or income). We have found that our accuracy in doing this, and the resulting width of the bounds for mobility, is significantly better when we are able to use retrospective information on the demographic composition of the household, the ownership of consumer durables and about the basic materials of housing. Such variables are typically collected only concurrently and not retrospectively in most household surveys. Nevertheless, they are not very time-consuming to collect retrospectively, and it is certainly much less costly to insert a few questions of this nature into surveys than to create a new panel. Since the results of this paper suggest the approach we propose offers promise, it seems worth experimenting with the inclusion of such questions in some upcoming nationally representative surveys in order to be able to provide basic estimates of poverty transitions.

[9] This work is being carried out by the World Bank's Latin American and the Caribbean office, not the authors of this study.

References

Antman, Francisca and David McKenzie (2007) "Earnings Mobility and Measurement Error: A Pseudo-Panel Approach", *Economic Development and Cultural Change* 56(1): 125-162.

Berney, L.R. and D.B. Blane (1997) "Collecting Retrospective Data: Accuracy of recall after 50 years judged against historical records", *Social Science and Medicine* 45(10): 1519-25.

Bourguignon, F. Goh C. and Kim, D. (2004) "Estimating Individual Vulnerability to Poverty with Pseudo-Panel Data", World Bank Policy Research Working Paper No. 3375, August, 2004.

Chen, S. and Ravallion, M. (2010) "The Developing World is Poorer than we Thought, But no Less Successful in the Fight Against Poverty", *Quarterly Journal of Economics*¸125(4): 1577-1625.

Deaton, Angus (1985) "Panel data from time series of cross-sections", *Journal of Econometrics* 30: 109-216.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2009) "Who are the microenterprise owners? Evidence from Sri Lanka on Tokman v. de Soto", forthcoming in Joshua Lerner and Antoinette Schoar (eds.) *International Differences in Entrepreneurship*. NBER, Cambridge, MA.

Demombynes, G., Elbers, C., Lanjouw, J., Lanjouw, P., Mistiaen, J. and Ozler, B. (2004) 'Producing a Better Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries'. In Shorrocks, A. and van der Hoeven, R. (eds) *Growth, Inequality and Poverty* (Oxford University Press).

Elbers, C., Lanjouw, J.O, and Lanjouw, P. (2002) "Micro-Level Estimation of Welfare" Policy Research Working Paper 2911, DECRG, The World Bank.

Elbers, C. Lanjouw, J.O. and Lanjouw, P. (2003) "Micro-level Estimation of Poverty and Inequality" *Econometrica*, 71(1): 355-364.

Elbers, C. Lanjouw, P. and Leite, P. (2010) 'Brazil Within Brazil: Testing the Poverty Map Methodology in Minas Gerais', mimeo, DECRG, the World Bank.

Fields, Gary, Robert Duval-Hernández, Samuel Freije Rodríguez, and María Laura Sánchez Puerta. (2007). "Earnings Mobility in Argentina, Mexico, and Venezuela: Testing the Divergence of Earnings and the Symmetry of Mobility Hypotheses." Mimeo. School of Industrial and Labor Relations, Cornell University.

Gibson, John (2001) "Measuring Chronic Poverty Without a Panel", *Journal of Development Economics* 65(2): 243-66.

Glewwe, Paul (2005) "How Much of Observed Mobility is Measurement Error? A Method to Reduce Measurement Error Bias, with an Application to Vietnam", Mimeo. University of Minnesota.

McKenzie, David (2005) "Measuring inequality with asset indicators", *Journal of Population Economics* 18:229-260.

Smith, James P. and Duncan Thomas (2003) "Remembrance of Things Past: Test-retest reliability of retrospective migration histories", *Journal of the Royal Statistical Society Series A*, 166(1): 23-49.

Verbeek, Marno (2008) "Pseudo-panels and repeated cross-sections", pp.369-383 in L. Matyas and P. Sevestre (eds.) *The Econometrics of Panel Data*. Springer-Verlag: Berlin.

Appendix 1

VIETNAM: Consumption models for upper bound estimates, Basic and Full models, VLSS:

Dependent var: Log Welfare 1992	Basic	Full
Age head in 1992	0.004***	.004***
	(0.001)	0.001
Rural birthplace	0.241***	0.053*
	(0.036)	(0.029)
Primary education or less in 1992	-0.147***	-.069**
	(0.029)	(0.029)
Female head of household	0.181***	.128**
	(0.062)	(0.059)
Head ever attended school and is over 30	0.186***	.081**
	(0.036)	(0.034)
Christian	-0.017	-0.052
	(0.049)	(0.039)
Kinh	-0.191***	-0.011
	(0.040)	(0.035)
Head ever attended school and is over 30 and is female	-0.001	-0.093
	(0.073)	(0.063)
Urban household in 1992		-0.056
		(0.105)
HH Head over 30 and urban home in 1992		0.107
		(0.106)
Census: Avg. # of persons > 60 in cluster		-5.044***
		(0.657)
Census: Avg. share of electrified homes in cluster		0.145***
		(0.053)
Census: Avg. rate of literacy in cluster		0.692***
		(0.124)
Census: Avg. share Buddhists in cluster		0.380***
		(0.107)
Census: Avg. share of homes with TV in cluster		0.088
		(0.094)
Head education level 1		0.210***
		(0.056)
Head education level 2		.234***
		(0.057)
Head education level 3		.318***
		(0.067)
Head education level 4		.302***
		(0.070)
Head education level 5		.241**
		(0.116)
Head works in agriculture		0.030
		(0.028)
Head is a professional		0.044
		(0.030)
Head operates self-employed business		.067***
		(0.021)

Household size		-.142***
		(0.020)
Sq. of household size		.006***
		(0.002)
Head is married		0.038
		(0.042)
HH owns black & white TV		.205***
		(0.031)
HH owns color TV		.388***
		(0.045)
HH owns motorbike		.280***
		(0.040)
HH owns electric fan		.137***
		(0.031)
Home has permanent flooring material		.105***
		(0.023)
Constant	7.130***	6.911***
	(0.045)	(0.132)
R-squared	0.133	0.533
N	1597	1584

Full model includes regional dummies.
* p< 0.10, ** p<0.05, *** p<0.01

INDONESIA: Consumption models for upper bound estimates, Basic and Full models, IFLS:

Age head	0.006***	0.000
	(0.001)	(0.001)
Educ head if over 25	0.050***	0.026**
	(0.005)	(0.011)
Literate	0.001	0.101**
	(0.044)	(0.049)
Head is male	-0.235***	-0.010
	(0.050)	(0.046)
Head birth place is small town	0.018	-0.051
	(0.040)	(0.036)
Head birth place is big city	0.153**	0.047
	(0.065)	(0.061)
Head birth place is other	-0.078	-0.216
	(0.245)	(0.346)
Urban		-0.088**
		(0.038)
Avg. rate of electrification in cmty		0.002***
		(0.001)
Village main road type is stone		-0.031
		(0.039)
Village main road type is dirt		0.054
		(0.059)
Cmty has a primary school		0.042
		(0.054)
Head self-employed		0.091*
		(0.052)
Head government worker		0.061
		(0.065)
Head works in private industry		0.034
		(0.056)
Head is unpaid family member		0.241**
		(0.118)
HH Farms		0.013
		(0.034)
Household size		-0.272***
		(0.029)
Sq. of household size		0.015***
		(0.003)
# of children < 5		-0.105***
		(0.023)
Log of Housing Floor Space, m2		0.198***
		(0.024)
Cook fuel: gas		0.746***
		(0.228)
Cook fuel: kerosene stove		0.271
		(0.218)
Cook fuel: firewood		-0.025
		(0.218)

Own a private car (0/1)		0.088***
		(0.029)
Have electric lighting in home (0/1)		0.055
		(0.049)
Constant	11.919	11.584
	(0.093)	(0.440)
R-squared	0.1969	0.417
N	2404	2304

Basic model in column 1 includes fixed effects for language spoken at home, education level of head's father, and household religion. Full model includes all of basic model, plus province of residence, head's education level, and primary roof material.
* p< 0.10, ** p<0.05, *** p<0.01