

Intra-generational Mobility and Repeated Cross-Sections

A Three-country Validation Exercise

Guillermo Cruces

Peter Lanjouw

Leonardo Lucchetti

Elizaveta Perova

Renos Vakis

Mariana Viollaz

The World Bank
Latin America and the Caribbean Region
Poverty, Equity and Gender Unit
December 2011



Abstract

This paper validates a recently proposed method to estimate intra-generational mobility through repeated cross-sectional surveys. The technique allows the creation of a “synthetic panel”—done by predicting future or past household income using a set of simple modeling and error structure assumptions—and thus permits the estimation of lower and upper bounds on directional mobility measures. The authors validate the approach in three different settings where good panel data also exist (Chile, Nicaragua, and Peru). In doing so, they also carry

out a number of refinements to the validation procedure. The results are broadly encouraging: the methodology performs well in all three settings, especially in cases where richer model specifications can be estimated. The technique does equally well in predicting short and long-term mobility patterns and is robust to a broad set of additional “stress” and sensitivity tests. Overall, the paper lends support to the application of this approach to settings where panel data are absent.

This paper is a product of the Poverty, Equity and Gender Unit, Latin America and the Caribbean Region. 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 rvakis@worldbank.org.

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

**Intra-generational Mobility and Repeated Cross-Sections:
A Three-country Validation Exercise***

Guillermo Cruces
CEDLAS-CONICET,
Universidad Nacional de La Plata

Peter Lanjouw
The World Bank

Leonardo Lucchetti
The World Bank

Elizaveta Perova
The World Bank

Renos Vakis
The World Bank

Mariana Viollaz
CEDLAS-CONICET,
Universidad Nacional de La Plata

Keywords: Poverty; Mobility; Pseudo-panels; Synthetic panels; Chile; Nicaragua; Peru.

JEL classification: O15, I32.

Sector Board: POV

* We are grateful to Monserrat Bustelo, Hai-Anh Dang, participants at the Institut d'Analisi Economica/World Bank workshop on "Socio-Economic Mobility and the Middle Class in Latin America" as well as participants at the World Bank "Summer Initiative on Research on Poverty Inequality and Gender" whose suggestions greatly improved earlier drafts of the paper. All remaining errors are ours. The views expressed here are those of the authors and should not be attributed to World Bank or its affiliates.

1. Introduction

Recent years have seen the study of intra-generational mobility increasingly capture the attention of policy makers and researchers. The type of policies needed to attack persistent poverty may be quite different from those required to address transient poverty or movements in higher parts of the income distribution. When measuring mobility, it is desirable to work with panel datasets that follow individuals or households over time. Unfortunately, such surveys pose substantial empirical challenges. First, because they are typically costly and complex to administer panel datasets (especially in developing countries) that track individuals or households over time are rare. This is particularly true for longer term panel data (that track the same unit of analysis for more than 5-10 year). Second, and connected to the previous point, it is usually complicated to revisit households who move physically or dropout from panel data surveys. As such, non-random attrition may significantly bias results, leading to an underestimation of the actual mobility in the general population (Antman and McKenzie 2007). Finally, classical measurement error will also introduce bias in the mobility estimates.

Because of growing concern to assess evaluate transitions into and out of poverty, an emerging body of research exists to develop techniques to overcome the major limitations of panel data sets by using cross-sectional surveys. Most of the literature has focused on what is commonly referred to as a pseudo-panel approach, which tracks cohorts of individuals over several periods of time. Recent developments on pseudo-panel analysis include Bourguignon, Goh, and Kim (2004) and Artman and McKenzie (2007). However, these studies usually impose significant data demands and structural assumptions in order to yield mobility measures out of repeated cross-sectional surveys. For example, more than two cross sections are often needed and specific functional forms have to be assumed for earnings dynamics (Lanjouw, Louto, and

McKenzie 2011). Another critique of the approach is that by aggregating average trends for a given group (or cohort) it assumes away key intra-group mobility, which may be equally or more important than aggregate mobility. A recent approach developed by Lanjouw et al. (2011) explores movements in and out of poverty by imposing fewer restrictions than earlier literature on pseudo-panels. The method produces lower and upper bound estimates of mobility which are expected to sandwich *true* mobility estimates obtained from actual panel datasets.

This paper seeks to validate this “synthetic panel” approach by estimating intra-generational mobility by means of repeated cross-sectional surveys. We focus on Chile, Nicaragua, and Peru; three countries for which we have panel datasets that follow households for more than two rounds. This allows us to compare true estimates of mobility using the three panel datasets against mobility estimates from applying the Lanjouw et al. (2011) synthetic panel approach in which individual rounds of the panels are treated as though they were repeated cross-sectional surveys.¹

In doing so, the paper makes four contributions to the existing literature. First, the study performs a range of sensitivity analyses and robustness tests to help validate the methodology proposed by Lanjouw et al. (2011) and probe its underlying assumptions.

Second, we improve on Lanjouw et al.’s (2011) inability to use retrospective information to improve the estimates. In their application to Vietnam and Indonesian data, Lanjouw et al (2011) force cross-sectional information to be time-invariant and retrospective by making use of the two-year panel structure of the dataset – i.e., they insert the value of variables from the first round into the second round of the panel. This provides an “ideal” but not entirely realistic context within which to apply the method. By contrast, our surveys include the information

¹ The use of “true” mobility in this paper relates to how close the approach comes to replicate panel data results. It remains that the typical panel data set may itself suffer from measurement error and high attrition rates, resulting in error-prone estimates of mobility.

needed to create retrospective asset ownership. We illustrate that these data carry us a considerable distance in the direction of mimicking the “ideal” set-up explored in Lanjouw et al (2011), thereby strengthening the practical appeal of this synthetic panel approach.

Third, our diverse settings and panel data allow us to test the performance of different interval lengths between the two rounds of cross-sectional surveys, from a one-year interval in Peru to an interval of ten-years in Chile. This is a key question for mobility analysis as it is important to know whether the technique can provide accurate predictions of mobility both in the short-term as well as the long-term. In fact, and as mentioned above, while short-term panel data are relatively common in developing country settings, they rarely exceed more than three years in length, making *any* analysis of long term mobility almost impossible. Evaluating the performance of this approach for long term mobility is therefore crucial.

Finally, Lanjouw et al. (2011) include only variables in levels as regressors in the underlying models of consumption. This paper introduces a variety of interactions between time invariant household characteristics and geographical controls and regional fixed effects (suggested by Lanjouw et al. 2011, but not applied). We find that the introduction of such interactions indeed increases the predictive power of models and thereby narrows the bound estimates yielded by the approach.

In all, our results indicate that the methodology performs well in predicting a range of mobility measures by means of two rounds of cross-sectional data; *true* mobility lies within the two bounds in most of the cases studied. We find that the particular specification of the underlying model of income/consumption matters for reducing the bounds, with considerable progress possible following the introduction of interaction terms and retrospective asset

ownership variables as regressors. Results are generally robust to a large number of sensitivity analyses and robustness checks.

The next section summarizes the technique. Section 3 discusses the data and the approach to ensure comparability of the tests across the three countries. The main results are presented in Section 4, while Section 5 discusses a range of additional robustness checks.

2. Methodology

This section largely relies on Lanjouw et al. (2011). We assume two rounds of repeated cross-sectional surveys. Calling y_{it} round t household log per capita consumption or income (where $t = 1, 2$) of household i and z the poverty line, we are interested in estimating the fraction of poor households in the first round of the survey who escaped poverty ($\Pr(y_{i2} > z | y_{i1} < z)$) or remain poor ($\Pr(y_{i2} < z | y_{i1} < z)$) in the second round of the survey, and the fraction of non-poor households in the first round of the survey who became poor ($\Pr(y_{i2} < z | y_{i1} > z)$) or remained non-poor ($\Pr(y_{i2} > z | y_{i1} > z)$) in the second round of the survey. This task cannot be performed directly by using repeated cross-sectional surveys, since all households are interviewed only once, either in the first or the second round of the survey.

However, we can straightforwardly estimate the relationship between income and time invariant characteristics in each round:

$$(1) \quad y_{it} = \beta_t' x_{it} + \varepsilon_{it} \quad t = 1, 2$$

where x_{it} is a vector of time-invariant characteristics (or characteristics that can be easily recalled from one round to the other one) of household i in round t of the survey and ε_{it} is an error term.

Using observations from the second round, we can predict consumption in the first round (\hat{y}_{i1}^2) by means of the same observed vector of time-invariant or retrospective characteristics (x_{i1}^2) and the first round OLS estimates of parameters $\hat{\beta}_1$, where the superscript refers to observations of

households surveyed in the second round.² Lower and upper bound estimates of mobility are derived from two different sets of assumptions about the correlation between the error term in the first round and in the second round.

Lanjouw et al. (2011) argue that the correlation between both error terms is likely to be non-negative.³ If we are willing to assume zero correlation between the first round and the second round error terms, Lanjouw et al. (2011) propose to predict consumption in the first round by randomly drawing with replacement for each household i in the second round from the empirical distribution of first round estimated residuals (denoted by $\tilde{\varepsilon}_{i1}^2$) as follows:

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

Equation 2 allows us to compute estimates of movements in and out of poverty. For example, the fraction of poor households in the first round who escaped poverty in the second time is given by:

$$(3) \quad \Pr(y_{2i}^2 > z | \hat{y}_{i1}^{2U} < z)$$

Since we are randomly drawing from the empirical distribution of estimated errors, we need to repeat the procedure R times and take average of equation 3 in order to estimate movements in and out of poverty.⁴

In all likelihood, however, the correlation between error terms will be positive. By assuming no correlation, equation 3 will provide an *upper bound* estimate of the mobility in and out of poverty. Lanjouw et al. (2011) propose estimating a *lower bound* on mobility by now assuming a perfect positive correlation between error terms. In this case estimates of residuals

² Section (5) discusses the robustness of changing the forecasting direction.

³ Correlation between error terms will be non-zero in two cases: (i) the error term includes an individual fixed effect and (ii) shocks to consumption persist over time. Lanjouw et al. (2011) argue that correlation between error terms will almost certainly be positive if the condition (ii) holds. In their study using Vietnamese and Indonesian data they present empirical support in favor of this assumption.

⁴ We replicate the procedure 50 times in this paper ($R=50$). Section (5) discusses the robustness of replicating the procedure 150 and 300 times.

from the second round ($\hat{\varepsilon}_{i2}^2$) can be directly used to predict consumption in the first round as follows

$$(4) \quad \hat{y}_{i1}^{2L} = \hat{\beta}_1' x_{i1}^2 + \hat{\varepsilon}_{i2}^2$$

Equation 4 allows us to compute lower bound estimates of movements in and out of poverty. For example, the fraction of poor households in the first round who escaped poverty in the second time is given by

$$(5) \quad \Pr(y_{2i}^2 > z | \hat{y}_{i1}^{2L} < z)$$

Since we are not drawing from the empirical distribution of estimated errors, we do not need to repeat the procedure R times as in the upper bound approach. In fact, this last approach provides a clean under-estimate of true mobility since we are using household-specific error terms (from the second round in this example). In other words, because mobility is estimated across two survey rounds in which the same disturbance term applies to both consumption measures, the lower-bound measure of mobility has been “purged” of classical measurement error and thereby provides a lower-bound estimated of “true” mobility.

3. Data and harmonizing the approach across countries

In order to validate the technique, we use three panel datasets for Chile, Nicaragua, and Peru. True panel estimates of mobility from each country are compared with mobility estimates that result from applying the Lanjouw et al. (2011) approach. The three panel datasets we use are: (i) the 1996, 2001, and 2006 CASEN panel survey from Chile; (ii) the 1998, 2001, and 2005 National Household Survey on Living Standards Measurement (EMNV in Spanish) panel survey from Nicaragua; and (iii) the 2004-2006 and the 2008-2009 Peruvian National Household Survey (ENAH0 in Spanish) panel surveys in Peru.

Our use of these three sets of surveys offers a number of advantages. First, the years considered in the analysis coincide with a period of sustained income growth and reduction of poverty and inequality in the Latin America region; mobility in and/or out of poverty is expected to be large. Second, having several years of panel datasets allows us to validate the technique for different lengths in time both *within* and *across* countries ranging from five to ten years in Chile, from three to eight years in Nicaragua and from one to two years in Peru. Some additional information on each dataset follows.

3.1 Chile (CASEN Survey)

We use the 1996, 2001, and 2006 CASEN Panel survey. The CASEN survey is carried out jointly by the Foundation for Overcoming Poverty (FSP), Ministry of Planning (Mideplan) and Social Observatory of the Universidad Alberto Hurtado (OSUAH) and its main objective is to study poverty dynamics and vulnerability.

The first round of the CASEN panel survey interviewed 20,948 individuals in the Third, Seventh, Eighth, and Metropolitan regions of Chile, representing approximately 60 percent of total population. The survey was carried out between November and December 1996 and it has information mainly on education, employment, income, health, labor history, participation, and housing. The second round of the CASEN survey was conducted between November and December 2001 and surveyed 18,851 individuals; from which 15,038 were interviewed in 1996 (the corresponding attrition rate is 28.2 percent). Finally, the last round of the survey was performed between November and December 2006 and January and February 2007. This round interviewed 14,996, from which 10,287 were surveyed in 1996. The last two rounds of the survey provide the same information as the CASEN 1996 survey.

3.2. Nicaragua (ENMV Survey)

To study mobility in and out of poverty in Nicaragua we use the 1998, 2001, and 2005 rounds of the EMNV panel survey. The survey was developed by the National Institute of Statistics and Censuses with the technical and financial assistance of the World Bank, the United Nations Development Programme (UNDP), the Inter-American Development Bank (IDB), and the government of Nicaragua. The main objective of the EMNV survey is to study the socio-economic characteristics and the living conditions of the population of Nicaragua.

The first round of the EMNV panel survey interviewed 4,209 households and has national coverage. The survey was fielded between May and July 1998 and provides information on family relationships, education, health, economic activity, time, housing, consumption, household enterprise, and agro-pastoral activities. The second round of the EMNV was fielded between May and July 2001, while the last round was fielded between July and October 2005.

3.3. Peru (ENAHO Survey)

In order to estimate poverty mobility in Peru we use the 2008 and 2009 ENAHO survey,⁵ which was developed by the Peruvian Statistics Bureau (INEI). The ENAHO is a nationally representative survey yielding rich information on education, employment, income and expenditure, health, participation, social programs, housing, and perceptions. The survey's main objectives are to measure poverty evolution and households' living conditions.

The first round of the panel survey interviewed 7,560 households from January to December 2008, while the second round of the survey interviewed 7,546 households from January to December 2009.

⁵ Section 5 discusses the use of the 2004, 2005, and 2006 ENAHO panel survey to test the robustness of results to different panel lengths.

3.4. Harmonizing the approach across Countries

In order to better assess the synthetic panel technique as it is applied simultaneously in the three different countries and settings we apply a number of data harmonizing protocols. First, to avoid potential bias from using panel data to conduct the tests, we follow Lanjouw et al. (2011) in by splitting every panel dataset in every country into two randomly drawn sub-samples and then treating one sub-sample from each round as two repeated cross-sectional surveys. We then use these two repeated cross-sections to estimate mobility by applying the method described in section 2. These results are then compared with the true panel consisting of the other sub-sample.

Second, we apply the same specifications across countries (to the extent possible based on data availability). Specification 1 is the most restrictive in that it only uses variables that best adhere to the time invariance assumptions: household head age, age squared, gender, years of education, and ethnicity. Specification 2 adds geographical controls and regional fixed effects while the specification 3 adds more flexible interaction terms between the first two specifications. Finally, specification 4 adds retrospective asset ownership information (Peru only).

Third, for each country we predict household income using time invariant characteristics from round 1 and the returns to those characteristics from round 0. This yields a synthetic panel that uses the actual welfare measure from round 1 and the simulated one from round 0. We repeat the procedure in the reverse order as an additional validation exercise.

In addition, we apply household weights to address survey design and explore the use of alternative weighting schemes (no weights or individual weights in the robustness tests). For the upper bound mobility calculation, we apply 50 replications to estimate consumption in period 0 (and we also explore a higher number of replications as another check). Finally, we restrict our

analysis to households whose head is between 25 and 65 years of age in order to avoid life cycle effects which can invalidate the time invariance assumption.

4. Main results

In this section we present the main results by estimating a wide range of mobility and welfare outcomes to test how well the technique performs.

4.1. Poverty and Directional Mobility Measures (transition matrices)

We start our analysis by first comparing true poverty with estimated poverty rates that arise from applying the method proposed in section 2. For each country, Table 1 reports point estimates and 95 percent confidence intervals of actual poverty rates in the fifth column, together with lower and upper bound predictions of poverty rates for the four specifications detailed in section 3.4.⁶ Columns one through four present lower bound estimates, while columns six through nine show upper bound estimates. The models overall work well; most of the predictions lie within the 95 percent confidence interval of the true poverty rate. This is true especially for specifications that include interaction terms and retrospective information. In addition, predictive power increases and the bounds narrow considerably when moving to more complex specifications, specially the inclusion of interaction terms and retrospective asset ownership.

We then proceed to estimate directional mobility as measured by changes in the proportion of households that move across poverty status. Tables 2a, 2b, and 2c provide estimates of poverty transitions in Chile, Nicaragua, and Peru, respectively. The tables show point estimates and 95 percent confidence intervals of true panel mobility, together with lower and upper bound predictions of poverty mobility for all the specifications explained in section

⁶ The models used for the upper and lower bound estimates are presented in the Appendix. We use official poverty lines in each country.

3.4. The predictions again work well; true estimates lie within the estimated bounds. For example, the model that includes interaction terms suggests that between 11.09 and 21.78 percent of households escape poverty in Chile between the first and the second round of the survey, while the actual panel dataset suggests that actual upward mobility is somewhere between 16.87 and 22.29 percent. Similarly, true downward mobility lies between 1.79 and 4.11 percent in Chile, while the model with interactions predicts that between 2.25 and 5.13 entered poverty between first and second rounds of the survey.

As before, predicted power increases rapidly when moving from the first specification towards the fourth specification; bounds are also reduced considerably when increasing the number of variables. This is especially the case for the specifications that include interaction terms and retrospective variables. These results suggest that the use of the most complex model available is recommended.

As expected, the correlation of disturbance terms between the first and second rounds is always positive. The results also show that, as in the case of Lanjouw et al. (2011), this residual correlation declines when moving from the first toward the fourth specification; presumably because the additional explanatory variables are better able to capture the effect of shocks and fixed effects. For example, the model that includes regional controls and fixed effects reduces the residual correlation from 0.69 to 0.64 in Peru. Correlation is then slightly reduced to 0.62 when we include interaction terms and further reduced to 0.56 when we include retrospective variables.

4.2. Directional Mobility as a Percentage of Income Change – Non-anonymous Growth Incidence Curves (GIC)

Figures 1a, 1b, and 1c draw on the synthetic panel procedure to present estimated non-anonymous GICs. Non-anonymous GICs plot income/consumption growth against ventiles of the initial distribution as in Bourguignon (2010). Predictions are based on the third specification explained in section 3.4 and use parameter estimates obtained from the whole sample. Lower bound estimates are characterized by the gray dashed line, while upper bound estimates are represented by the black dashed line. Predictions are compared to the true panel GIC based on panel data analysis, which is characterized by the solid black line.

The figures show a greater income/consumption growth for the lower ventiles of the initial income/consumption distribution in all three countries. The true panel GIC lies within the estimated GIC bounds. Interestingly, the GIC based on the lower bound seems to do a slightly better job to capturing the panel GIC for most of the distribution.

4.3. Sub-population Directional Mobility Measures – Changes in the Proportion of Specific Groups that move across poverty status

Tables 2a, 2b, and 2c refer to overall mobility in Chile, Nicaragua, and Peru. One relevant question is whether the analysis performs well in predicting mobility for specific sub-groups, for example urban and rural areas separately. We experiment with a number of groups, which also vary from country to country depending on data availability. These include: region of residence; gender and education of the household head; urban/rural sector; occupation; sector of occupation; household ownership; access to water and electricity; and ethnicity.

Figures 3a, 3b, and 3c present true panel point estimates, as well as lower and upper bound estimates, for different population sub-groups in Chile, Nicaragua, and Peru. Each figure

is divided into four panels which show: (i) the proportion of poor households in the first round who remain poor in the second round, (ii) the proportion of poor households who escape poverty in the second round, (iii) the proportion of non-poor households who enter poverty in the second round, and (iv) the proportion of non-poor household who remain non-poor in the second round.

Predictions are based on the third specification explained in section 3.4 and use parameter estimates obtained using the whole sample. Lower bounds are characterized by black dots, while upper bound estimates are represented by gray dots. Predictions of mobility are compared to the true mobility based on panel data analysis, which is symbolized by the 45-degree line. The closer the dots are to the 45-degree line the better the model predicts actual mobility.

The technique performs well for almost all the sub-groups. As expected, upper bound estimates tend to overstate movements in and out of poverty, while the opposite happens to lower bound estimates. Dots are generally close to the line, and the true value lies between the lower and upper bound estimates for most of the sub-groups. Lower bound estimates seem to perform slightly better for Peru and Nicaragua, while the opposite applies to Chile.

5. Robustness checks

This section performs a range of sensitivity analysis and robustness tests to the application of the synthetic panel procedure we implemented in the preceding section.

5.1. Sensitivity Analysis Using Different Poverty Lines

Except for the GICs, all previous estimations were computed using the official poverty lines of each country. In this section we test whether the method is sensitive to the specific poverty line selected. Figures 3a, 3b, and 3c show true point estimates, as well as lower and upper bound

estimates, for different poverty lines in Chile, Nicaragua, and Peru.⁷ Lower and upper bounds are characterized by gray and black dashed lines, respectively. Predictions of mobility are compared to the true panel mobility based on panel data analysis, which is described by solid black lines. Mobility estimates are computed for alternative poverty lines from \$2 to \$15 (USD at 2005 PPP a day), except for Nicaragua where poverty lines go up to \$7 since there is no household with income above that threshold. Each figure is divided into four panels as in the preceding subsection.

All solid lines have the expected shape for the three countries. The proportion of households who remain poor in both rounds increases from zero to 100 when the poverty line also increases, while the opposite happens with the proportion of households who are never poor. In addition, the proportion of the population who enters or escapes poverty shows an invert U-shape. In general, true panel estimates fall within bounds in all three countries. This is relevant since results suggest that the cutoff point selected is not critical for the analysis, meaning that the technique can be used to explore transitions at various parts of the distribution (e.g., poverty, middle class, etc.). However, it is not clear whether lower or upper bound estimates perform better in terms of low or high values of poverty lines. A possible explanation for this differential performance could be the change in residual correlation along the income distribution.

5.2. Performance of Short vs. Long Panel

The three countries we study provide different types of panel lengths: one and two year spans in Peru, three, five and seven years span in Nicaragua, and five and ten years in Chile. Table 3 shows mobility estimates for the range of panel lengths we can calculate to test whether this affects the results. Panel A shows two equally spanned periods for Chile: from 1996 to 2001 in

⁷ Unless otherwise noted, all of the robustness checks in section 5 use predictions that are based on the third specification explained in section 3.4.

columns one through three and from 2001 to 2006 in columns four through six. Panel B shows two periods of different length for Nicaragua: from 1998 to 2001 in columns one through three and from 1998 to 2005 in columns four through six. Lastly, panel C also shows two periods of different length for Peru: from 2004 to 2005 in columns one through three and from 2004 to 2006 in columns four through six.

Results suggest that the technique performs well irrespective of the panel length. For Nicaragua, the results weakly suggest that moving from shorter to longer panel lengths reduces the bounds. For example, model estimates suggest that between 21.23 and 32.69 percent of households remain poor between 1998 and 2001, while the actual panel dataset suggests that poverty persistence is somewhere between 24.44 and 31.27 percent. True poverty persistence is similar between 1998 and 2005, between 23.43 and 30.68 percent. However, bound estimates are narrower; between 21.16 and 26.57 percent of households remain poor between 1998 and 2005.

These findings show that the method is of great value for mobility analysis. As we mentioned earlier, panel data length rarely exceeds more than three years in developing country settings. These results show that the method provides accurate prediction of mobility in both the short-term as well as long-term.

5.3. Choice of Welfare Measure: Income vs. Consumption

Table 4 tests the robustness of results from using different welfare measure for the analysis. Specifically, we test whether using log of income instead of log of consumption as dependent variables affects the estimates. Results are available only for Peru and Nicaragua due to data availability. Results for Nicaragua are shown in panel A, while results for Peru are presented in panel B. The table shows that the method performs well irrespective of the welfare measure

employed. Upper and lower bound estimates sandwich true mobility confidence interval as before.

5.4. Alternative Specifications

We also perform numerous empirical exercises to test the robustness of the findings to the explanatory variables used in the underlying models.

5.4.1. Actual Variables from Panel vs. those Created Using Cross Sections

Columns four and nine of table 2c present estimates of mobility in Peru based on underlying models of consumption that include retrospective asset ownership as regressors (derived from questions in the cross sectional data on length of ownership of specific assets). In the case of Lanjouw et al. (2011), the authors force cross-sectional information to be time-invariant and retrospective by inserting the panel survey value of variables from the first round into the second round.

To test how well our “constructed” retrospective variables are, we force variables to be retrospective as in Lanjouw et al. (2011) and compare them with those in table 2c. Panel A of table 5 shows these results. Once again, the methodology is robust to the use of actual data; upper and lower bound estimates sandwich true mobility confidence interval and results are similar to those in table 2c. These results have important implications since they suggest that constructing retrospective variables should be done whenever they are available in surveys.⁸

5.4.2. Place of Residence

Columns three and eight of table 2c shows estimates of mobility in Peru based on a consumption model that includes department of birth as regressor, as well as their interactions with household time invariant characteristics. Arguably, department of birth is of great value since it is time-

⁸ We also do a simple test of significance between the 6 asset variables we use in the analysis to test whether the panel values from the first round are not statistically different from those constructed using the information in round two. We find no significant difference for 5 out of the 6 variables.

invariant by definition, a necessary requirement for the method to work properly. Panel B of table 5 tests the robustness of results by replacing department of birth fixed effects by current place of residence fixed effects. Results show that the methodology works well when using current place of residence instead of retrospective data on place of birth; 95 percent true panel confidence interval of mobility lies within upper and lower bound estimates and results are similar to those in table 2c.

5.4.3. Sub-national Aggregates: Census vs. Household Survey

Columns three and eight of table 2c presents estimates of mobility in Peru based on underlying models of consumption that include sub-national controls from the 2005 Census at the village-level (i.e., percentage of female head of households, percentage of head of households with primary education incomplete, percentage of head of households with secondary education incomplete, percentage of head of households with secondary education complete, percentage of households with electricity, and percentage of households with access to water), as well as their interaction with household time invariant characteristics. As in the case of department of birth, sub-national controls are of great value since they are time invariant. Panel C of table 5 tests the robustness of results by replacing Census variables at the village-level by potentially Census variables at the department level. Predictions are based on the third specification explained in section 3.4. The technique performs well when using Census variables at the department-level, 95 percent true panel confidence interval of mobility lies within upper and lower bound estimates. Results are also robust to the use of department-level controls from the survey instead of the Census (not shown). Taken together, these results suggest that we can directly use survey information to construct sub-national controls and interaction terms to increase precision of estimates.

5.5. Additional Tests

5.5.1. Forecasting Direction

In section 4, we estimated models of income/consumption using one random half of the first round of the survey and predict mobility in the second random half of the first round of the survey using observations from the second round. Upper and lower bound estimates are then compared to the true mobility of the second half of the first round of the survey. Table 6 tests the robustness of results changing the forecasting direction. Panel A shows results for Chile, panel B for Nicaragua, and panel C for Peru. Predictions are based on the third specification explained in section 3.4. Results suggest that the technique performs well irrespective of the forecasting direction.⁹

5.5.2. Number of Replications for Upper Bound Estimates

Upper bound estimates in all preceding analysis have used 50 replications. Table 7 tests the robustness of results to different number of repetitions. Column three and four show upper bound estimates using 150 and 300 repetitions, respectively. First column shows lower bound estimates, while the second column shows true estimates. Since we do not need to repeat any process, lower bound and true estimates are identical to those in tables 2a, 2b, and 2c. Panel A shows results for Chile, panel B for Nicaragua, and panel C for Peru. Results are robust to the number of repetitions; upper bound estimates in columns three and four are almost identical to those in tables 2a, 2b, and 2c. In this sense, this suggests that the precision gains beyond 50 replications are limited.

⁹ True mobility in table 6 does not coincide with true mobility in table 3 because there is a change in weights between first and second round of surveys. Results are the same when we estimate mobility without using sampling weights (results not shown).

5.5.3. *Survey Design*

All previous results are weighted using household sampling weights. Table 8 tests the robustness of results to different survey design. Columns one through three show estimates using no-weights, while columns four through six use individual sampling weights (household sampling weights times household size). Panel A shows results for Chile, panel B for Nicaragua, and panel C for Peru. Predictions are based on the third specification explained in section 3.4. Results suggest that the technique performs well irrespective of survey design. True mobility confidence interval lies within or overlaps upper and lower bound estimates of mobility irrespective of the weights used.

8. Conclusion

Recently, there has been growing interest among policy makers and researchers to study intra-generational mobility in and out of poverty. Panel datasets constitute the most appropriate information source for the study of mobility. Unfortunately, such surveys are rare and, where they exist, generally span only a few years and follow a relatively small number of households or individuals over time. To overcome this limitation, there is also a growing literature which studies intra-generational mobility by means of cross-sectional surveys via the application of “pseudo-panel” methods. A new method proposed by Lanjouw et al. (2011), which largely relies on insights from poverty-mapping techniques (Elbers et al, 2003), yields lower and upper bound estimates of mobility using cross sectional surveys. The significant advantage of this “synthetic panel” method is that it imposes fewer restrictions and structural assumptions than the earlier literature on pseudo-panels.

This paper uses three panel datasets for Chile, Nicaragua, and Peru to validate the methodology proposed by Lanjouw et al. (2011). By providing a wide range of new sensitivity analyses and robustness checks, this paper shows further that the technique is remarkably flexible with respect to the model specification and the choice of variables. Results indicate that the methodology performs well in predicting actual mobility in and out of poverty by means of two rounds of cross-sectional data; true mobility lies within the two bounds most of the time. Specification of the underlying model of income/consumption matters for reducing the bounds, which can be narrowed considerably via the introduction of interaction terms and asset ownership as controls in the underlying model of consumption/income. More important, the method appears equally well-suited to the estimation of short term mobility as long term mobility.

References

- Antman, Francisca, and David McKenzie. 2007. "Earnings mobility and measurement error: a pseudo-panel approach." *Economic Development and Cultural Change*, 56(1): 125-161.
- Bourguignon, Francois, Chor-ching Goh, and Dae Il Kim. 2004. "Estimating individual vulnerability to poverty with pseudo-panel data." World Bank Policy Research Working Paper No. 3375.
- Bourguignon, Francois. 2010. "Non-anonymous growth incidence curves, income mobility and social welfare dominance." *Journal of Economic Inequality*, DOI: 10.1007/s10888-010-9159-7.
- ENAH0. 2004-2006 and 2008-2009. National Household Surveys. Lima, Peru. Peruvian Statistics Bureau (INEI).
- Lanjouw, Peter, Jill Louto, and David McKenzie. 2011. "Using Repeated Cross-Sections to Explore Movements in and out of Poverty." World Bank Policy Research Working Paper 5550.
- Ravallion, Martin & Chen, Shaohua, 2003. "Measuring pro-poor growth," *Economics Letters*, 78(1):93-99

Tables and Figures

Table 1: Poverty Headcount - Synthetic vs. Actual Panel Data

Status in Year[1]	Lower Bound Estimates				Truth [5]	Upper Bound Estimates			
	[1]	[2]	[3]	[4]		[6]	[7]	[8]	[9]
<i>Panel A: Chile 1996</i>									
Poverty Rate	15.16	17.02	16.44	.	24.22	28.34	23.20	24.24	.
					(20.07, 28.37)				
<i>Panel B: Nicaragua 2001</i>									
Poverty Rate	40.63	39.44	40.65	.	39.03	43.56	41.01	41.25	.
					(34.08, 43.97)				
<i>Panel C: Peru 2008</i>									
Poverty Rate	35.16	34.22	33.57	33.19	33.35	36.52	34.88	34.81	34.97
					(30.44, 36.25)				
Obs. Panel A	823	823	823	.	823	823	823	823	.
Obs. Panel B	684	684	684	.	684	684	684	684	.
Obs. Panel C	2,368	2,368	2,368	2,368	2,368	2,368	2,368	2,368	2,368

Data source: Chile CASEN 1996 and 2006; Nicaragua EMNV 2001 and 2005; and Peru ENAHO 2008 and 2009.

Note: R-squared is calculated for opposite halves of the total Year[1] sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Columns [1] and [6] show a simple model with household time invariant characteristics. Columns [2] and [7] add sub-national controls from census and region of birth fixed-effects. Columns [3] and [8] add interactions between household time invariant characteristics and sub-national controls. Columns [4] and [9] add retrospective asset ownership. Results in column [5] show actual panel poverty. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 2a: Transition Matrices - Synthetic vs. Actual Panel Data
Chile 1996 and 2006

Status in 1996, 2006	Lower Bound Estimates			Truth	Upper Bound Estimates		
	[1]	[2]	[3]		[4]	[5]	[6]
Poor, Poor	7.46	6.68	5.35	4.64 (3.19, 6.07)	2.96	2.65	2.46
Poor, Non-poor	7.70	10.35	11.09	19.59 (16.87, 22.29)	25.39	20.55	21.78
Non-poor, Poor	0.13	0.92	2.25	2.96 (1.79, 4.11)	4.64	4.94	5.13
Non-poor, Non-poor	84.71	82.06	81.31	72.82 (69.77, 75.85)	67.02	71.85	70.63
R-squared	0.21	0.31	0.43	.	0.21	0.31	0.43
Residual Correlation	0.26	0.22	0.22	.	0.26	0.22	0.22
Observations	823	823	823	823	823	823	823

Data source: CASEN, 1996-2006.

Note: R-squared is calculated for opposite halves of the total 1996 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Columns [1] and [5] show a simple model with household time invariant characteristics. Columns [2] and [6] add sub-national controls and region fixed effects. Columns [3] and [7] add interactions between household time invariant characteristics and sub-national controls. Results in column [4] show actual panel mobility. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 2b: Transition Matrices - Synthetic vs. Actual Panel Data
Nicaragua 2001 and 2005

Status in 2001, 2005	Lower Bound Estimates			Truth	Upper Bound Estimates		
	[1]	[2]	[3]		[4]	[5]	[6]
Poor, Poor	40.63	39.44	37.34	35.68 (32.08, 39.26)	31.70	31.17	30.73
Poor, Non-poor	0.00	0.00	3.31	3.35 (2.00, 4.70)	11.86	9.84	10.52
Non-poor, Poor	21.17	22.36	24.46	26.12 (22.82, 29.41)	30.10	30.63	31.07
Non-poor, Non-poor	38.20	38.20	34.89	34.85 (31.27, 38.42)	26.34	28.36	27.68
R-squared	0.30	0.45	0.53	.	0.30	0.45	0.53
Residual Correlation	0.58	0.53	0.52	.	0.58	0.53	0.52
Observations	684	684	684	684	684	684	684

Data source: EMNV, 2001-2005.

Note: R-squared is calculated for opposite halves of the total 2001 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Columns [1] and [5] show a simple model with household time invariant characteristics. Columns [2] and [6] add sub-national controls from census and region fixed effects. Columns [3] and [7] add interactions between household time invariant characteristics and sub-national controls. Results in column [4] show actual panel mobility. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 2c: Transition Matrices - Synthetic vs. Actual Panel Data
Peru 2008 and 2009

Status in 2008, 2009	Lower Bound Estimates				Truth	Upper Bound Estimates			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Poor, Poor	33.90	31.72	28.70	29.01	23.61 (21.89, 25.31)	16.20	17.64	17.56	18.57
Poor, Non-poor	1.26	2.50	4.87	4.18	9.74 (8.54, 10.93)	20.32	17.24	17.25	16.40
Non-poor, Poor	0.19	2.38	5.40	5.09	10.49 (9.25, 11.72)	17.90	16.46	16.54	15.53
Non-poor, Non-poor	64.65	63.41	61.03	61.73	56.17 (54.16, 58.16)	45.59	48.67	48.66	49.51
R-squared	0.39	0.52	0.57	0.62	.	0.39	0.52	0.57	0.62
Residual Correlation	0.69	0.64	0.62	0.56	.	0.69	0.64	0.62	0.56
Observations	2,368	2,368	2,368	2,368	2,368	2,368	2,368	2,368	2,368

Data source: INEI - Peruvian National Household Survey (ENAHO), 2008-2009.

Note: R-squared is calculated for opposite halves of the total 2008 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Columns [1] and [6] show a simple model with household time invariant characteristics. Columns [2] and [7] add sub-national controls from census and region of birth fixed-effects. Columns [3] and [8] add interactions between household time invariant characteristics and sub-national controls. Columns [4] and [9] add retrospective variables (asset ownership). Results in column [5] show actual panel mobility. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 3: Transition Matrices - Synthetic vs. Actual Panel Data Using Different Panel Length

Status in Year[1], Year[2]	Panel Survey I			Panel Survey II		
	Lower Bound Estimates	Truth Confidence Interval	Upper Bound Estimates	Lower Bound Estimates	Truth Confidence Interval	Upper Bound Estimates
	[1]	[2]	[3]	[4]	[5]	[6]
<i>Panel A: Chile; Years in Panel Survey I: 1996-2001; Years in Panel Survey II: 2001-2006</i>						
Poor, Poor	10.43	(5.44, 8.17)	5.19	8.48	(3.18, 6.00)	3.49
Poor, Non-poor	7.37	(9.59, 13.02)	21.74	13.94	(8.21, 12.29)	25.46
Non-poor, Poor	2.95	(5.23, 7.91)	8.19	0.48	(2.99, 5.73)	5.47
Non-poor, Non-poor	79.24	(72.96, 77.63)	64.88	77.10	(78.13, 83.43)	65.58
<i>Panel B: Nicaragua; Years in Panel Survey I: 1998-2001; Years in Panel Survey II: 1998-2005</i>						
Poor, Poor	32.69	(24.44, 31.27)	21.23	26.57	(23.43, 30.68)	21.16
Poor, Non-poor	4.23	(4.95, 8.81)	14.54	1.27	(1.72, 4.58)	6.91
Non-poor, Poor	9.70	(11.84, 17.21)	21.16	32.91	(28.59, 36.24)	38.32
Non-poor, Non-poor	53.38	(46.91, 54.53)	43.08	39.25	(33.41, 41.31)	33.61
<i>Panel C: Peru; Years in Panel Survey I: 2004-2005; Years in Panel Survey II: 2004-2006</i>						
Poor, Poor	34.64	(28.68, 33.30)	23.24	34.35	(26.11, 30.73)	22.08
Poor, Non-poor	9.31	(9.34, 12.46)	21.21	10.64	(12.21, 15.76)	21.93
Non-poor, Poor	5.59	(7.79, 10.68)	17.00	4.07	(8.45, 11.52)	16.33
Non-poor, Non-poor	50.45	(46.36, 51.35)	38.56	50.94	(45.03, 50.14)	39.66
Obs. Panel A	1,313	1,313	1,313	852	852	852
Obs. Panel B	664	664	664	577	577	577
Obs. Panel C	1,543	1,543	1,543	1,468	1,468	1,468

Data source: Chile CASEN, 1996, 2001, and 2006; Nicaragua EMNV, 1998, 2001, and 2005; and Peru ENAHO 2004, 2005, and 2006.

Note: R-squared is calculated for opposite halves of the total Year[1] sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Controls are the same as in columns [3] and [7] of table 2c. Results in column [2], and [5] show actual panel mobility. 95 percent confidence interval between parentheses.

Table 4: Poverty Dynamics from - Synthetic vs. Actual Panel Data
Using Income as the Welfare Measure

Status in Year[1], Year[2]	Lower Bound Estimates [1]	Truth [2]	Upper Bound Estimates [3]
<i>Panel A: Nicaragua</i>			
Poor, Poor	41.66	(32.57, 39.77)	29.42
Poor, Non-poor	7.80	(6.10, 10.20)	17.35
Non-poor, Poor	12.16	(14.78, 20.50)	24.40
Non-poor, Non-poor	38.38	(34.38, 41.66)	28.83
<i>Panel B: Peru</i>			
Poor, Poor	27.06	(20.47, 23.81)	16.15
Poor, Non-poor	7.54	(11.89, 14.63)	20.23
Non-poor, Poor	3.74	(7.52, 9.78)	14.65
Non-poor, Non-poor	61.66	(53.93, 57.93)	48.97
Observations Panel A	684	684	684
Observations Panel B	2,368	2,368	2,368

Data source: Nicaragua EMNV 2001, and 2005; Peru Peruvian National Household Survey (ENAHO) 2008 and 2009.

Note: R-squared is calculated for opposite halves of the total 2008 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Results in columns [1] and [3] are based on a model with household time invariant characteristics, sub-national controls from census, and interactions between household time invariant characteristics and sub-national controls. Panel A shows results for Nicaragua, while panel B shows results for Peru+A2. Results in column [2] show actual mobility. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 5: Transition Matrices - Synthetic vs. Actual Panel Data
Using Alternative Specifications, Peru 2008 and 2009

Status in 2008, 2009	Lower Bound Estimates [1]	Truth [2]	Upper Bound Estimates [3]
<i>Panel A: Using Actual Instead of Retrospective Data</i>			
Poor, Poor	29.18	(21.89, 25.31)	18.59
Poor, Non-poor	4.22	(8.54, 10.93)	16.64
Non-poor, Poor	4.92	(9.25, 11.72)	15.51
Non-poor, Non-poor	61.68	(54.16, 58.16)	49.27
<i>Panel B: Using Actual Department of Residence Instead of Department of Birth</i>			
Poor, Poor	29.46	(21.89, 25.31)	17.55
Poor, Non-poor	5.30	(8.54, 10.93)	16.55
Non-poor, Poor	4.63	(9.25, 11.72)	16.55
Non-poor, Non-poor	60.60	(54.16, 58.16)	49.36
<i>Panel C: Using Sub-national Controls from Census at the Departmental-level</i>			
Poor, Poor	28.49	(21.89, 25.31)	17.63
Poor, Non-poor	4.77	(8.54, 10.93)	16.43
Non-poor, Poor	5.60	(9.25, 11.72)	16.47
Non-poor, Non-poor	61.13	(54.16, 58.16)	49.47
Observations	2,368	2,368	2,368

Data source: INEI - Peruvian National Household Survey (ENAHO), 2008-2009.
Note: R-squared is calculated for opposite halves of the total 2008 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Results in columns [1] and [3] are based on a model with household time invariant characteristics, sub-national controls from census, and interactions between household time invariant characteristics and sub-national controls. Panel A adds asset ownership. Results in panel C are based on sub-national controls from census at the departmental-level (instead of at the village-level). Results in column [2] show actual panel mobility. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 6: Transition Matrices - Synthetic vs. Actual Panel Data
Changing the Forecasting Direction

Status in Year[2], Year[1]	Lower Bound Estimates [1]	Truth Confidence Interval [2]	Upper Bound Estimates [3]
<i>Panel A: Chile 1996 and 2006</i>			
Poor, Poor	12.17	(3.72, 6.77)	6.28
Poor, Non-poor	4.88	(2.21, 4.71)	12.95
Non-poor, Poor	11.93	(16.17, 21.51)	17.81
Non-poor, Non-poor	71.03	(69.38, 75.49)	62.95
<i>Panel B: Nicaragua 2001 and 2005</i>			
Poor, Poor	35.46	(32.08, 39.27)	29.53
Poor, Non-poor	20.17	(22.83, 29.42)	31.00
Non-poor, Poor	3.57	(2.00, 4.70)	9.50
Non-poor, Non-poor	40.80	(31.28, 38.42)	29.97
<i>Panel C: Peru 2008 and 2009</i>			
Poor, Poor	28.33	(21.64, 25.05)	16.59
Poor, Non-poor	4.21	(8.72, 11.13)	17.55
Non-poor, Poor	4.35	(8.16, 10.51)	16.09
Non-poor, Non-poor	63.11	(55.40, 59.39)	49.76
Observations in Panel A	823	823	823
Observations in Panel B	684	684	684
Observations in Panel C	2,364	2,364	2,364

Data source: Chile CASEN 1996 and 2006, Nicaragua EMNV 2001 and 2005, and Peru ENAHO 2008 and 2009.

Note: R-squared is calculated for opposite halves of the total Year[2] sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Results in columns [1] and [3] are based on a model with household time invariant characteristics, sub-national controls from census, and interactions between household time invariant characteristics and sub-national controls. Results in column [2] show actual panel mobility. 95 percent confidence interval between parentheses. Upper bound estimations are based on 50 repetitions.

Table 7: Transition Matrices - Synthetic vs. Actual Panel Data
Using Different Number of Replications for the Upper Bound Estimates

Status in Year[1], Year[2]	Lower Bound Estimates	Truth Confidence Interval	Upper Bound Estimates	
			[3]	[4]
<i>Panel A: Chile 1996 and 2006</i>				
Poor, Poor	5.35	(3.19, 6.07)	2.52	2.49
Poor, Non-poor	11.09	(16.87, 22.29)	21.50	21.64
Non-poor, Poor	2.25	(1.79, 4.11)	5.08	5.10
Non-poor, Non-poor	81.31	(69.77, 75.85)	70.91	70.76
<i>Panel B: Nicaragua 2001 and 2005</i>				
Poor, Poor	37.34	(32.08, 39.26)	30.76	30.77
Poor, Non-poor	3.31	(2.00, 4.70)	10.58	10.57
Non-poor, Poor	24.46	(22.82, 29.41)	31.04	31.03
Non-poor, Non-poor	34.89	(31.27, 38.42)	27.62	27.63
<i>Panel C: Peru 2008 and 2009</i>				
Poor, Poor	28.70	(21.89, 25.31)	17.56	17.57
Poor, Non-poor	4.87	(8.544, 10.93)	17.21	17.21
Non-poor, Poor	5.40	(9.251, 11.72)	16.53	16.52
Non-poor, Non-poor	61.03	(54.16, 58.16)	48.70	48.70
Number of Repetitions	.	.	150	300
Observations on Panel A	823	823	823	823
Observations on Panel B	684	684	684	684
Observations on Panel C	2,368	2,368	2,368	2,368

Data source: Chile CASEN, 1996-2006; Nicaragua EMNV 2001-2005; and Peru ENAHO 2008-2009.

Note: R-squared is calculated for opposite halves of the total 1996 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Controls are the same as in columns [3] and [7] of tables 2a, 2b, and 2c. Results in column [3] are based on 150 repetitions, while results in column [4] are based on 300 repetitions. Results in column [2] show actual panel mobility. 95 percent confidence interval between parentheses.

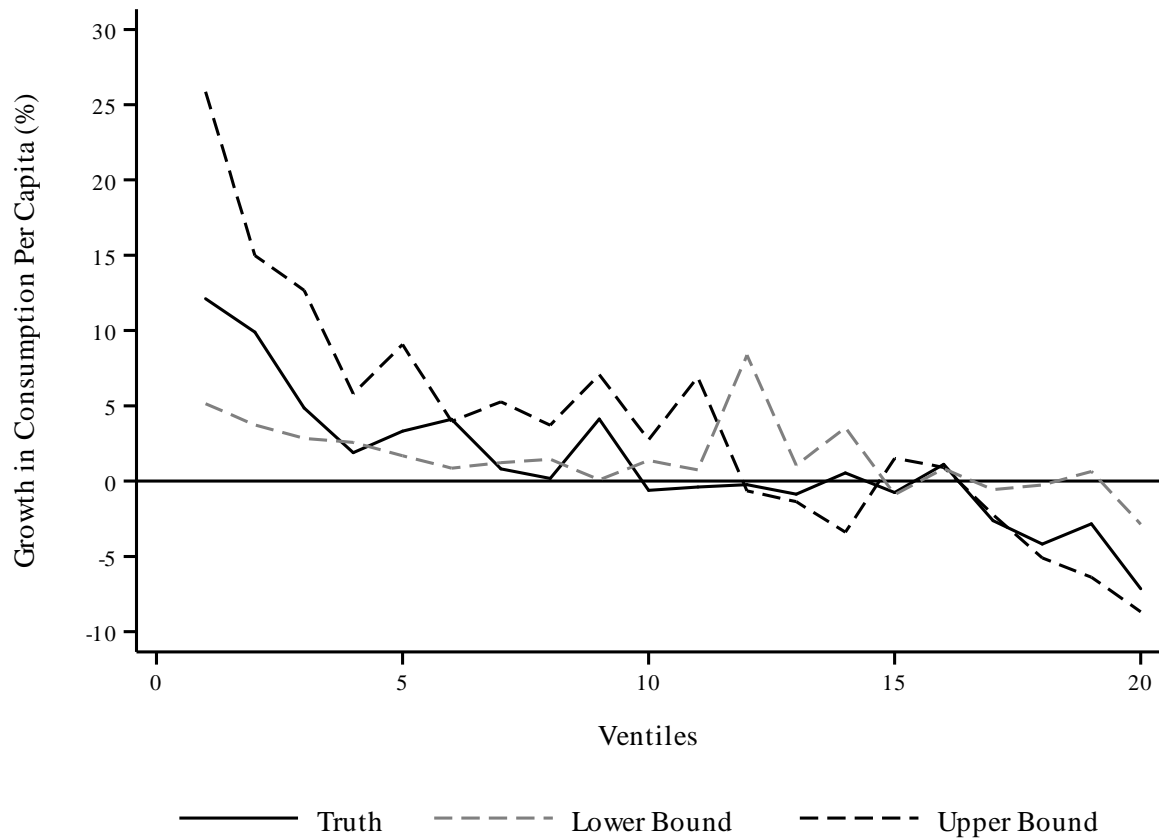
Table 8: Transition Matrices - Synthetic vs. Actual Panel Data
Using Different Survey Designs

Status in Year[1], Year[2]	Estimates Without Using Surveying-weights			Estimates Using Individual-level Surveying-weights		
	Lower Bound	Truth Confidence Interval	Upper Bound	Lower Bound	Truth Confidence Interval	Upper Bound
	[1]	[2]	[3]	[4]	[5]	[6]
<i>Panel A: Chile 1996 and 2006</i>						
P, P	15.80	(9.35, 13.72)	8.15	6.59	(4.39, 7.64)	3.04
P, NP	22.84	(24.05, 30.13)	29.61	11.43	(19.38, 25.07)	21.00
NP, P	0.73	(3.49, 6.46)	8.38	2.84	(2.17, 4.65)	6.39
NP, NP	60.63	(52.98, 59.76)	53.86	79.14	(65.15, 71.51)	69.57
<i>Panel B: Nicaragua 2001 and 2005</i>						
P, P	42.69	(37.53, 44.91)	34.65	43.03	(38.41, 45.82)	34.46
P, NP	2.49	(2.01, 4.71)	10.68	2.44	(1.05, 3.23)	8.20
NP, P	22.95	(21.19, 27.63)	30.99	26.51	(24.07, 30.76)	35.08
NP, NP	31.87	(27.52, 34.46)	23.68	28.02	(24.93, 31.69)	22.26
<i>Panel C: Peru 2008 and 2009</i>						
P, P	31.12	(24.12, 27.65)	18.81	34.46	(26.83, 30.47)	20.56
P, NP	4.65	(8.87, 11.30)	16.48	4.92	(8.47, 10.85)	15.64
NP, P	5.49	(9.47, 11.97)	17.80	6.14	(10.6, 13.25)	20.03
NP, NP	58.74	(51.28, 55.30)	46.91	54.48	(47.72, 51.74)	43.76
Obs. in Panel A	823	823	823	823	823	823
Obs. in Panel B	684	684	684	684	684	684
Obs. in Panel C	2,368	2,368	2,368	2,368	2,368	2,368

Data source: Chile CASEN, 1996-2006; Nicaragua EMNV 2001-2005; and Peru ENAHO 2008-2009.

Note: R-squared is calculated for opposite halves of the total Year[1] sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results in columns [1], [2], and [3] are not weighted. Results in columns [4], [5], and [6] are weighted using individual-level survey-sampling weights. Controls are the same as in columns [3] and [7] of tables 2a, 2b, and 2c. Results in column [2] and [5] show actual panel mobility. 95 percent confidence interval between parentheses. P, P refers to Poor, Poor; P, NP refers to Poor, Non-Poor; NP, P refers to Non-poor, Poor; and NP, NP refers to Non-poor, Non-poor

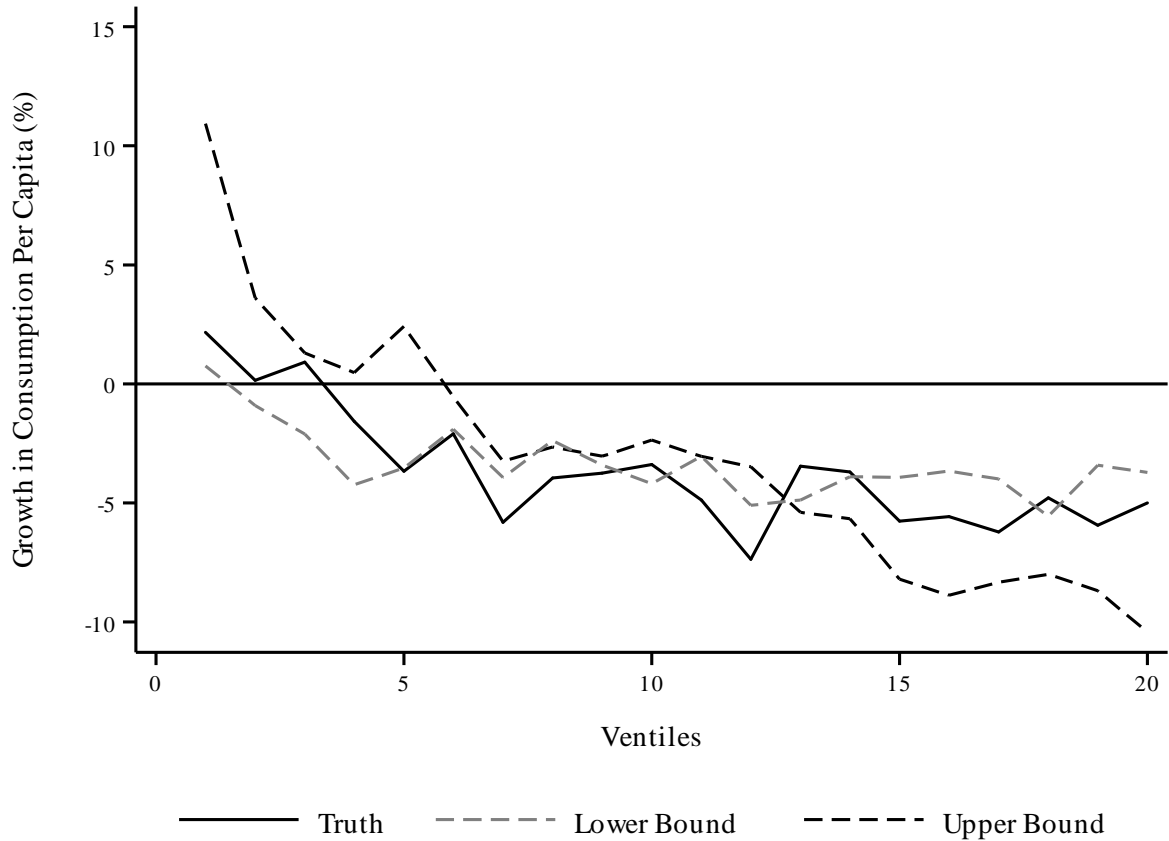
Figure 1a: Non-anonymous Growth Incidence Curve - Synthetic vs. Actual Panel Data
Chile 1996 and 2006



Data source: CASEN, 1996 - 2006.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Upper bound estimations are based on 50 repetitions.

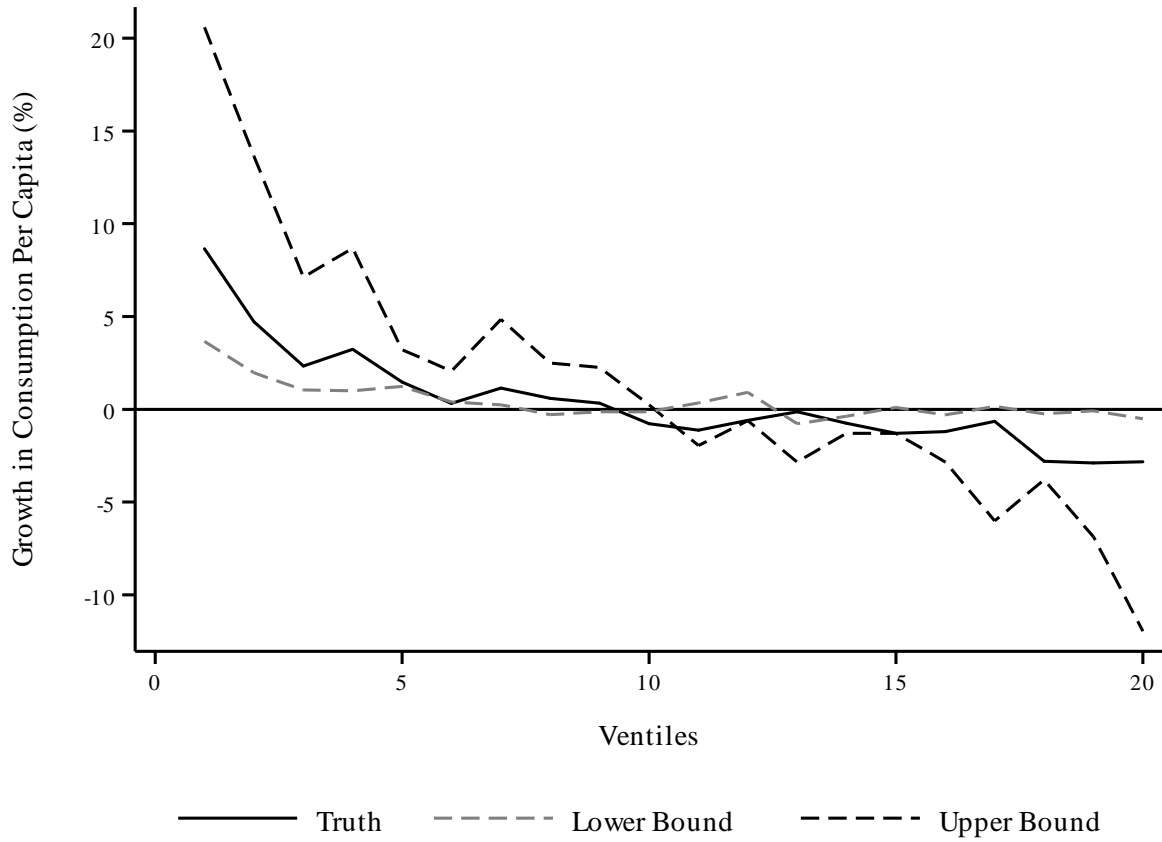
Figure 1b: Non-anonymous Growth Incidence Curve - Synthetic vs. Actual Panel Data
Nicaragua 2001 - 2005



Data source: EMNV, 2001 - 2005.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Upper bound estimations are based on 50 repetitions.

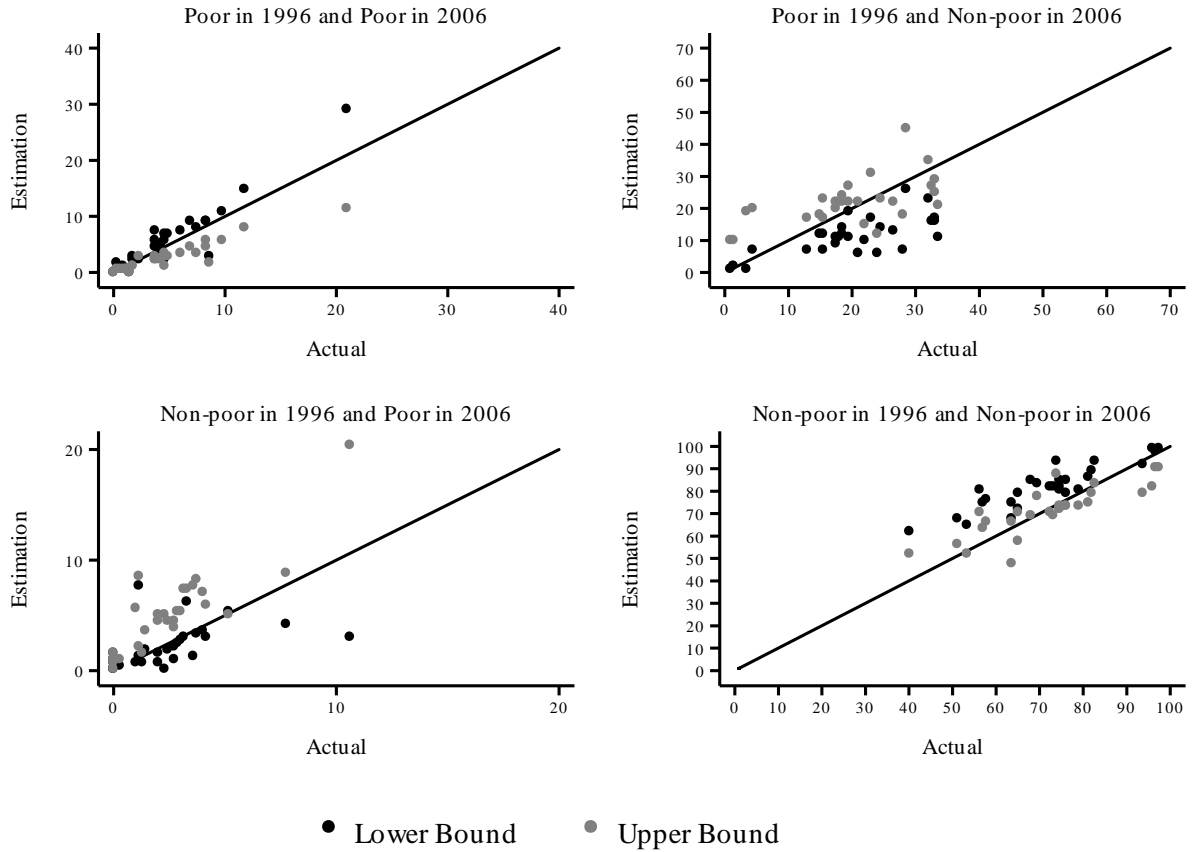
Figure 1c: Non-anonymous Growth Incidence Curve - Synthetic vs. Actual Panel Data
Peru 2008 and 2009



Data source: INEI - Peruvian National Household Survey (ENAHO), 2008-2009.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Upper bound estimations are based on 50 repetitions.

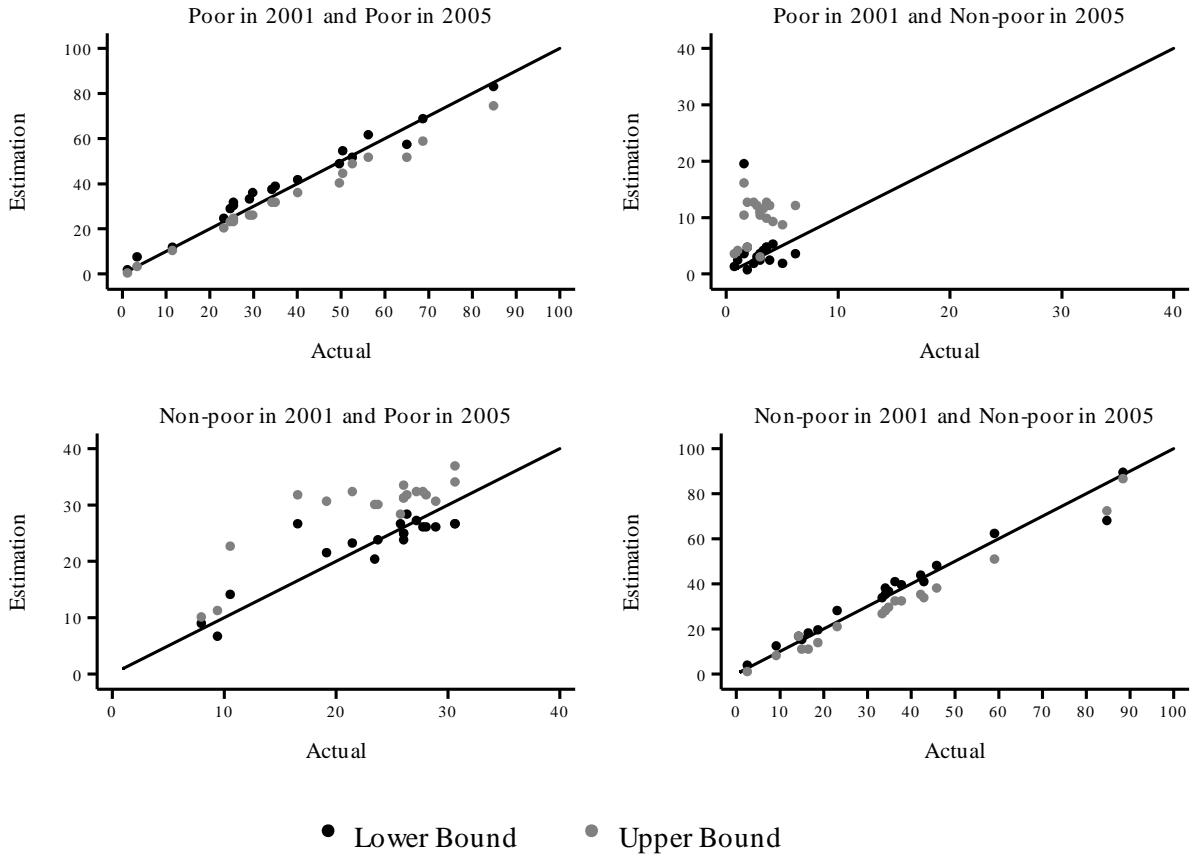
Figure 2a: Poverty Dynamics - Synthetic vs. Actual Panel Data
by Sub-population Groups, Chile 1996 and 2006



Data source: CASEN, 1996-2006.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. The 45 degree line shows actual panel mobility. Upper bound estimations are based on 50 repetitions.

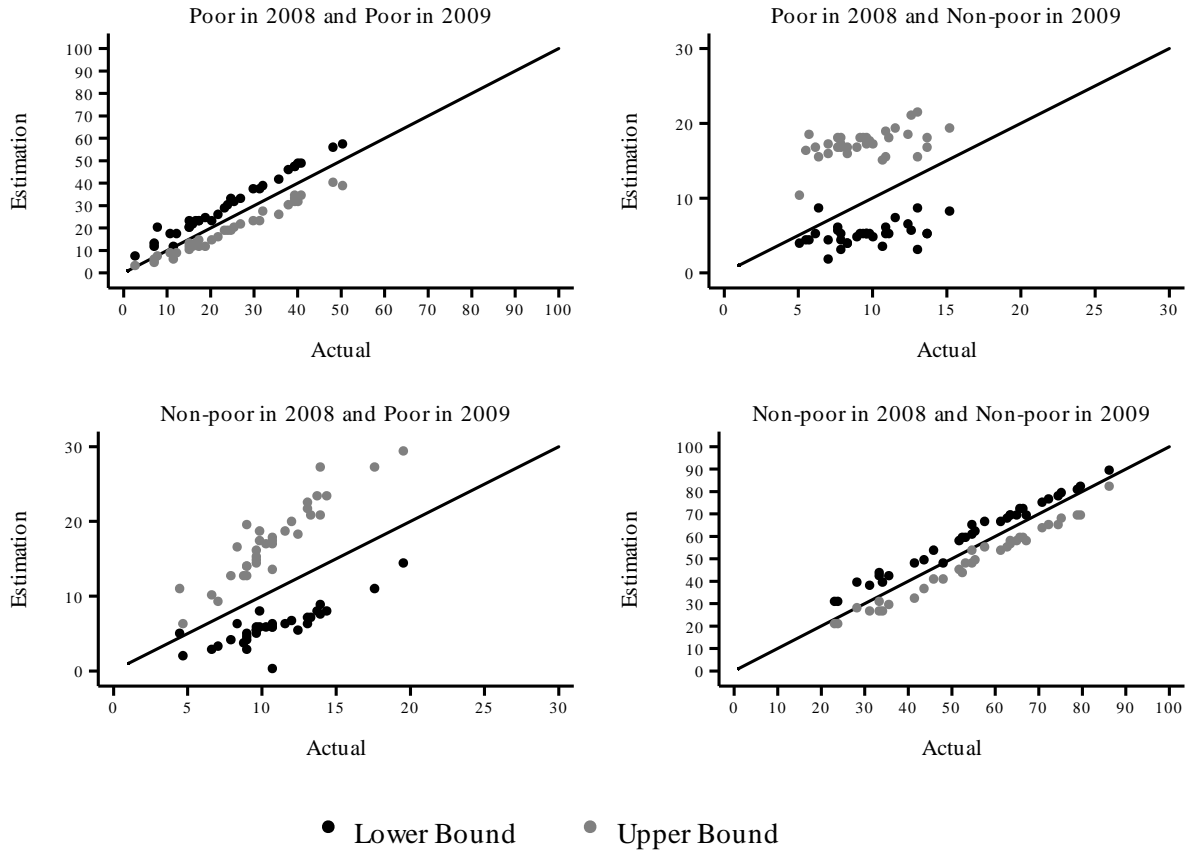
Figure 2b: Poverty Dynamics - Synthetic vs. Actual Panel Data
by Sub-population Groups, Nicaragua 2001 - 2005



Data source: EMNV, 2001 - 2005

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. The 45 degree line shows actual panel mobility. Upper bound estimations are based on 50 repetitions.

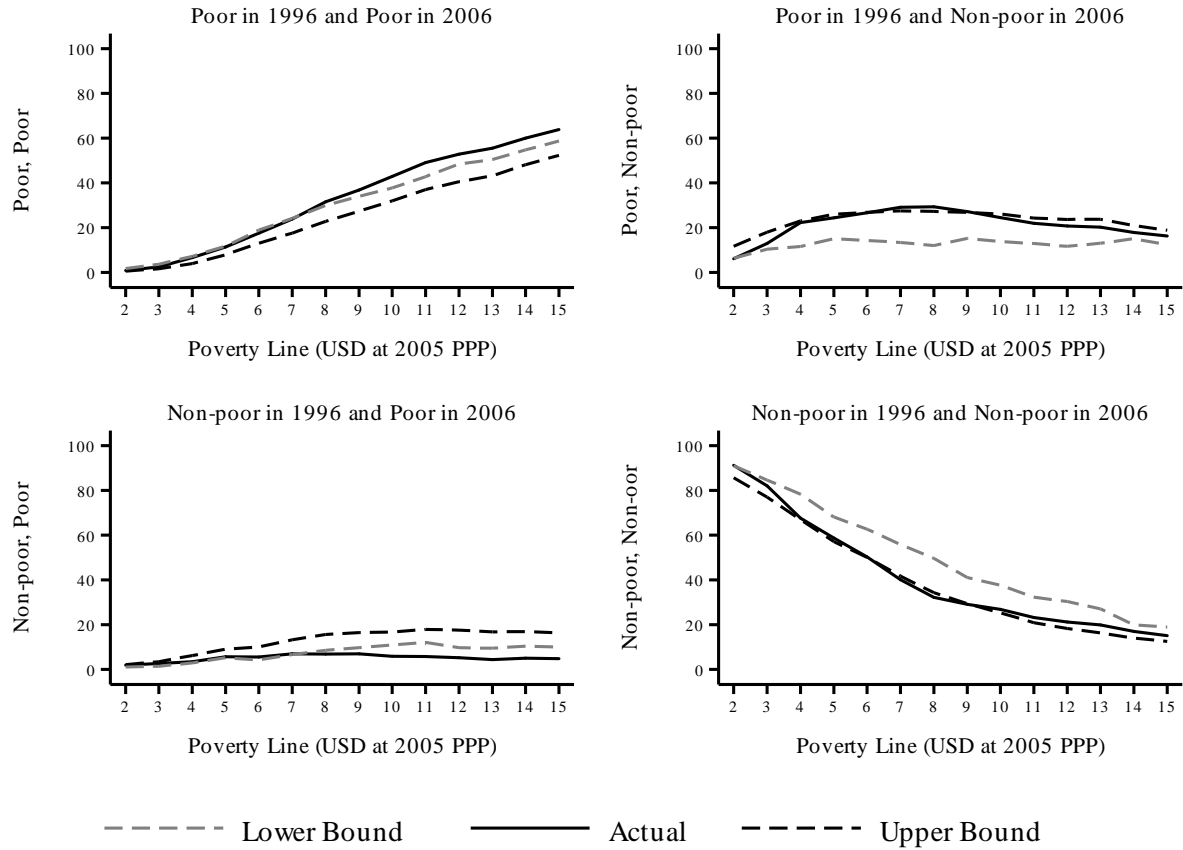
Figure 2c: Poverty Dynamics - Synthetic vs. Actual Panel Data
by Sub-population Groups, Peru 2008 and 2009



Data source: INEI - Peruvian National Household Survey (ENAHO), 2008-2009.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. The 45 degree line shows actual panel mobility. Upper bound estimations are based on 50 repetitions.

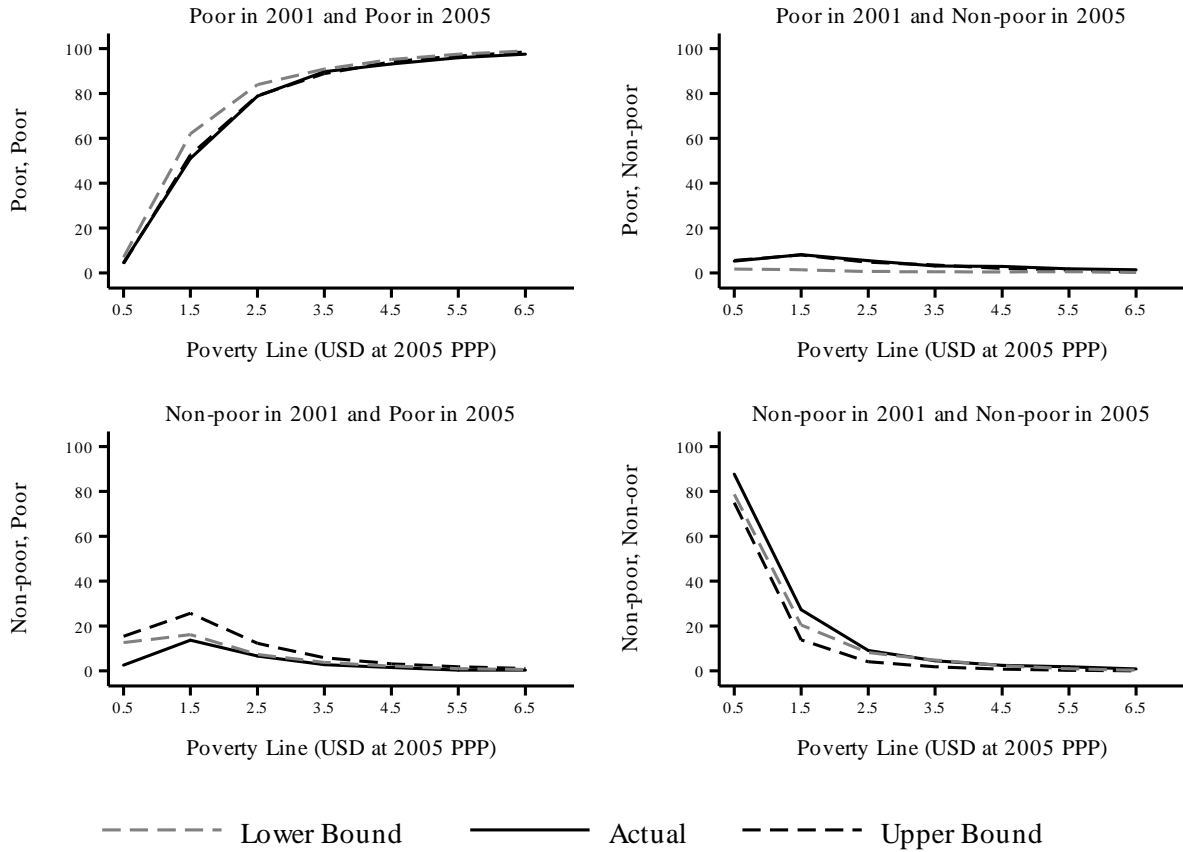
Figure 3a: Poverty Dynamics - Synthetic vs. Actual Panel Data for Alternative Poverty Lines, Chile 1996 and 2006



Data source: Chile CASEN 1996 and 2006.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Upper bound estimations are based on 50 repetitions.

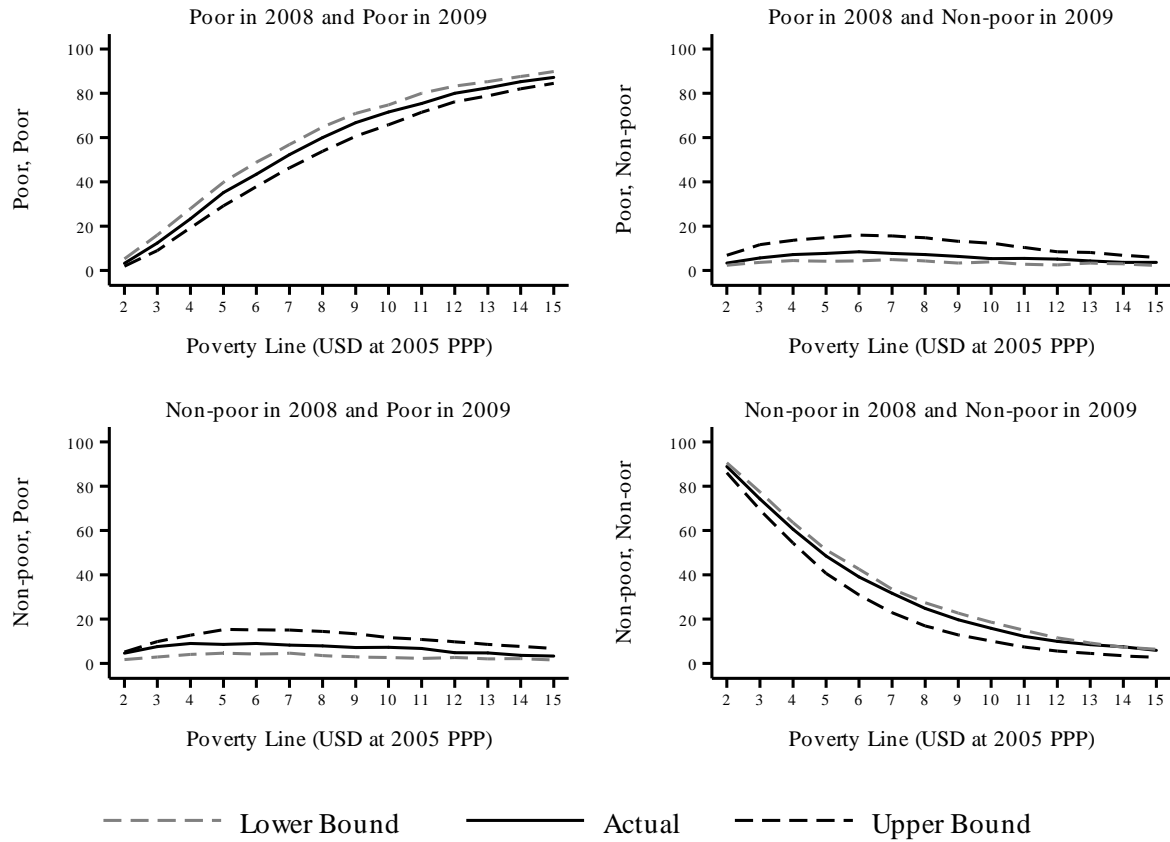
Figure 3b: Poverty Dynamics - Synthetic vs. Actual Panel Data for Alternative Poverty Lines, Nicaragua 2001 - 2005



Data source: EMNV 2001 and 2005.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Upper bound estimations are based on 50 repetitions.

Figure 3c: Poverty Dynamics - Synthetic vs. Actual Panel Data for Alternative Poverty Lines, Peru 2008 and 2009



Data source: INEI - Peruvian National Household Survey (ENAHO), 2008-2009.

Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Upper bound estimations are based on 50 repetitions.

Appendix

Table A1: Log Income Estimates
Chile 1996

Dependent Variable: Log Income	[1]	[2]	[3]
Male	-0.394 [0.297]	-0.369* [0.211]	-2.627 [1.637]
Age in 1996	0.033 [0.091]	0.025 [0.081]	1.918*** [0.599]
Age Squared in 1996	0.000 [0.001]	0.000 [0.001]	-0.022*** [0.007]
Years of Education	0.124*** [0.018]	0.102*** [0.017]	-0.026 [0.169]
% of Female Head of HHs		-1.057 [0.646]	4.578 [12.862]
% of Head of HHs With Incomplete Primary Education		-0.722 [1.048]	21.283 [15.864]
% of Head of HHs With Incomplete Secondary Education		-2.398** [1.015]	32.029** [13.693]
% of Head of HHs With Complete Secondary Education or More		-0.213 [1.205]	58.139*** [21.883]
Region Fixed Effects	Yes	Yes	Yes
Interaction Terms	No	No	Yes
Number of HHs	824	824	824

Data source: CASEN 1996.

Note: R-squared is calculated for opposite halves of the total 2008 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Column [1] shows a simple model with household time invariant characteristics. Column [2] adds sub-national controls from census and region of birth fixed-effects. Column [3] adds interactions between household time invariant characteristics and sub-national controls.

Table A2: Log Consumption Estimates
Nicaragua 2008

Dependent Variable: Log Consumption	[1]	[2]	[3]
Male	-0.078 [0.060]	0.035 [0.057]	-1.834 [2.955]
Age in 2008	-0.058** [0.025]	-0.059*** [0.022]	-0.075 [1.040]
Age Squared in 2008	0.001*** [0.000]	0.001*** [0.000]	0.003 [0.012]
Years of Education	0.093*** [0.007]	0.073*** [0.007]	0.450 [0.352]
Language: Spanish and Foreign	-0.155 [0.119]	-0.326 [0.210]	
% of Female Head of Households in 2001		0.924*** [0.338]	1.314 [7.619]
% of Head of HHs With Complete Primary Education in 2001		0.324 [0.351]	-0.854 [8.151]
% of Head of HHs With Complete Secondary Education in 2001		-0.461 [0.970]	14.252 [21.168]
% of Head of HHs With Complete Terciary in 2001		-0.148 [0.854]	-11.455 [16.747]
% of HHs with Access to Electricity in 2001		0.402* [0.209]	2.389 [4.683]
% of HHs with Access to Water in 2001		0.146 [0.172]	-2.883 [3.892]
% of HHs with Toilet in 2001		0.333 [0.252]	9.135* [5.382]
Average Household Size in 2001		-0.005 [0.051]	2.905*** [0.996]
Department of Birth Fixed Effects	Yes	Yes	Yes
Interaction Terms	No	Yes	Yes
Number of HHs	685	685	685

Data source: EMNV, 2001.

Note: R-squared is calculated for opposite halves of the total 2008 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Column [1] shows a simple model with household time invariant characteristics. Column [2] adds sub-national controls from census and region of birth fixed-effects. Column [3] adds interactions between household time invariant characteristics and sub-national controls.

Table A3: Log Consumption Estimates
Peru 2008

Dependent Variable: Log Consumption	[1]	[2]	[3]	[4]
Male	-0.206*** [0.038]	-0.100*** [0.036]	0.061 [0.550]	0.526 [0.550]
Age in 2008	-0.006 [0.013]	-0.013 [0.011]	0.179 [0.176]	0.167 [0.165]
Age Squared in 2008	0.000* [0.000]	0.000** [0.000]	-0.001 [0.002]	-0.001 [0.002]
Years of Education	0.088*** [0.003]	0.054*** [0.003]	0.133** [0.062]	0.087 [0.063]
Language: Spanish and Foreign	0.280*** [0.033]	0.198*** [0.041]	-1.221** [0.532]	-0.952* [0.517]
Percentage of Male Head of HHs		-0.796** [0.380]	-4.303 [7.286]	-7.198 [7.182]
% of Head of HHs With Incomplete Primary Education		-0.409* [0.210]	-0.873 [3.905]	-0.025 [3.638]
% of Head of HHs With Incomplete Secondary Education		1.209** [0.616]	2.477 [11.585]	4.068 [10.679]
% of Head of HHs With Complete Secondary Education or More		1.381*** [0.315]	-1.074 [5.437]	0.440 [5.284]
% of HHs with Access to Electricity		0.238*** [0.085]	0.248 [1.656]	0.738 [1.554]
% of HHs with Access to Water		0.105 [0.076]	2.605* [1.388]	1.955 [1.281]
Department of Birth Fixed Effects	Yes	Yes	Yes	Yes
Interaction Terms	No	No	Yes	Yes
Asset Ownership in 2008	No	No	No	Yes
Number of HHs	2,368	2,368	2,368	2,368

Data source: INEI - Peruvian National Household Survey (ENAHO), 2008.

Note: R-squared is calculated for opposite halves of the total 2008 sample. Results are constrained to the panel sample of HHs whose heads are between 25 and 65 years old. Results are weighted using household-level survey-sampling weights. Column [1] shows a simple model with household time invariant characteristics. Column [2] adds sub-national controls from census and region of birth fixed-effects. Column [3] adds interactions between household time invariant characteristics and sub-national controls. Column [4] adds retrospective variables (asset ownership).