Who Escaped Poverty and Who Was Left Behind?

A Non-Parametric Approach to Explore Welfare Dynamics
Using Cross-Sections

Leonardo Lucchetti

WORLD BANK GROUP

Poverty and Equity Global Practice Group

October 2017

Abstract

This paper proposes a non-parametric adaptation of a recently developed parametric technique to produce point estimates of intra-generational economic mobility in the absence of panel data sets that follow individuals over time. The method predicts past individual income or consumption using time-invariant observable characteristics, which allows the estimation of mobility into and out of poverty, as well as household-level income or consumption growth, from cross-sectional data. The paper validates this method by sampling repeated cross-sections out of actual panel

data sets from three countries in the Latin America region and comparing the technique with mobility from panels. Overall, the method performs well in the three settings; with few exceptions, all estimates fall within the 95 percent confidence intervals of the panel mobility. The quality of the estimates does not depend in general on the sophistication level of the underlying welfare model's specifications. The results are encouraging even for those specifications that include few time-invariant variables as regressors.

This paper is a product of the Poverty and Equity Global Practice 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 author may be contacted at llucchetti@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.

**Who Escaped Poverty and Who Was Left Behind?
A Non-Parametric Approach to Explore Welfare Dynamics Using Cross-
Sections**

Leonardo Lucchetti
The World Bank

Keywords: Poverty, Welfare dynamics; Poverty transitions; Synthetic panels.

JEL classification: O15, I32.

Sector Board: POV

I am grateful to Tanida Arayavechkit, Oscar Calvo, Andres Castaneda, Christoph Lakner, Moritz Meyer, Emmanuel Skoufias, and participants at a World Bank seminar whose suggestions greatly improved earlier drafts of the paper. All remaining errors are mine.
Email: llucchetti@worldbank.org

1. Introduction

Panel datasets that follow individuals and households over time are considered to be an excellent source of information for measuring intra-generational economic mobility. However, panels are often difficult to administer, costly, and not representative of the entire population of a country; they usually cover short periods of time; and suffer from non-random attrition that may bias mobility estimates. In addition, panels are only available in few countries, which restricts their use in regional analysis. Given all these limitations, panel datasets can impose serious constraints to the empirical analysis of the dynamics of poverty.

There are many alternatives to overcome the main limitations of panel datasets. “Synthetic Panels”, developed by [Dang et al. \(2014\)](#) and improved by [Dang and Lanjouw \(2013\)](#), is the most recent. [Dang et al. \(2014\)](#) implement a non-parametric approach to estimate an upper and a lower bound for mobility in and out of poverty at the household level as if we had actual panel datasets. The method estimates an income¹ model in the second round of cross-section data using a specification that includes only time-invariant covariates and retrospective regressors when available. Parameter estimates from this income model are also obtained in the first round of cross-sectional data and then applied to the same covariates in the second round. By doing this, we can predict the unobserved income in the first round of data for all households surveyed in the second round. Depending on how error terms are treated, the method generates an upper and a lower bound estimate of economic mobility. [Cruces et al. \(2015\)](#) perform a validation in three countries in Latin America and the Caribbean (LAC) where actual panel data are available: Chile, Nicaragua, and Peru. The authors show that the approach performs well in all three settings. [Ferreira et al. \(2012\)](#) apply the lower bound technique to predict upward economic mobility out of poverty in 18 countries in LAC between circa 1990 and 2010.

[Dang and Lanjouw \(2013\)](#) improve the methodology to obtain a point estimate—as opposed to lower and upper bounds—of poverty dynamics. The authors assume normality of the error terms and use the age-cohort correlation of residuals obtained from cross-sectional data in the first and second rounds to produce a point estimate of inter-generational poverty mobility. The authors validate the methodology using actual panel data from Bosnia-Herzegovina, Lao People’s

¹ For simplicity, I will refer to income, but consumption can also be used as the welfare measure.

Democratic Republic, Peru, United States, and Vietnam. [Vakis, Rigolini, and Lucchetti \(2016\)](#) apply this method to measure chronic poverty in 17 countries in LAC between circa 2004-2012.

The purpose of this paper is to introduce a non-parametric adaptation of the parametric point estimate developed by [Dang and Lanjouw \(2013\)](#). The method builds on the upper and lower bound mobility estimates developed by [Dang et al. \(2014\)](#). By taking a weighted average of the residuals obtained from estimations of the income models in the first and second rounds of data, we can predict a point estimate—as opposed to two bounds—of the unobserved income in the first round of data for all households surveyed in the second round. The paper validates this method by estimating intra-generational income mobility using cross-sectional data and by comparing it with true mobility estimates using actual panels in Chile, Nicaragua, and Peru.²

The proposal in this paper presents some strengths compared to mobility estimates computed by other studies in the literature. First, the key assumption required for the parametric approach developed by [Dang et al. \(2013, 2014\)](#) is the normality of the error terms in the underlying regressions. However, normality tests reject this assumption in Vietnam and Indonesia ([Dang et al. 2014](#)). By introducing a non-parametric approach, the method proposed in this paper does not assume any distribution of the error terms in the underlying income regressions.

Second, this non-parametric adaptation of the method predicts a point estimate of the unobserved income in the first round of data for all households surveyed in the second round, instead of just providing joint probabilities of mobility in and out of poverty as in the parametric approach. In doing so, the method allows estimating the income growth of every household between two rounds of surveys as if we had an actual household-level panel dataset.

Third, the paper performs a sensitivity test to study the robustness of results to changes in two underlying parameters—i.e., the number of repetitions that need to be run to obtain mobility estimates and the weights used for averaging residuals.

Fourth, validations in this paper are based on a panel that follows the same households for a period of five years. This allows us to validate the mobility estimates for periods of different lengths of time that include the same households. Validations are done for two periods; one that covers one year and another one that covers five years.

Finally, unlike previous studies, we use harmonized micro data that allow us to validate poverty dynamics using the same underlying income models and variables as the ones frequently

² Like in [Cruces et al. \(2015\)](#), the term “true” mobility refers to mobility estimates using actual panel data.

used in regional studies in LAC (such as [Cord et al. 2017](#); [Ferrerira et al. 2012](#); [Gasparini, Cicowiez, and Escudero 2013](#); and [Vakis et al. 2016](#)). This allow us to estimate income mobility for all LAC countries for which data are available using the exact same approach as the one proposed in this paper.

Results indicate that the methodology performs well in all the settings. In general, mobility estimates fall within the 95 percent confidence intervals of the true panel mobility. More importantly, the paper finds that a simple specification for the underlying income models that includes few time-invariant regressors is sufficient to obtain accurate estimates of mobility. Finally, the validation confirms that the method performs well at estimating household-level income growth (and not just join probabilities) between two rounds of surveys.

The next [section](#) summarizes the technique. [Section 3](#) discusses the data and the empirical approach. [Section 4](#) introduces the main validation results. [Section 5](#) presents mobility results for 17 Latin American countries. Finally, [Section 6](#) concludes.

2. Methodology³

Let y_{it} be round t household log per capita income (where $t=1, 2$) of household i and z the poverty line. The goal is to estimate the change in incomes between both periods ($\Delta y_i = y_{i2} - y_{i1}$) as well as all possible poverty dynamics: the proportion of poor households who escaped poverty ($\Pr(y_{i1} < z \text{ and } y_{i2} > z)$), who remained poor ($\Pr(y_{i1} < z \text{ and } y_{i2} < z)$), who became poor ($\Pr(y_{i1} > z \text{ and } y_{i2} < z)$), and who remained non-poor ($\Pr(y_{i1} > z \text{ and } y_{i2} > z)$).

The linear projection of income on a vector x_{it} of time-invariant characteristics (e.g., education of the household head under the assumption that the identity of the household head remains constant between rounds of data) of household i in round t is determined by

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

where ε_{it} is an error term. We can introduce superscripts to denote household observations surveyed in a given moment in time. Then, we can predict first round “unobserved” incomes of households surveyed in the second round ($\hat{y}_{i1}^2$) by multiplying their time-invariant characteristics (x_{i1}^2) and the first-round Ordinary Least Squares (OLS) estimates of parameters $\hat{\beta}_1$. We can obtain

³ This section largely relies on [Dang and Lanjouw \(2013\)](#), [Dang et al. \(2014\)](#), [Cruces et al. \(2015\)](#), and [Vakis et al. \(2016\)](#).

lower and upper bounds estimates of poverty mobility depending on the assumption that we consider with respect to the correlation between the error term in the first round and in the second round. [Dang et al. \(2014\)](#) explain that this correlation is likely to be non-negative. Upper bound estimates arise from assuming zero correlation. The authors suggest drawing randomly with replacement from the empirical distribution of first round estimated residuals (represented by $\hat{\varepsilon}_{i1}^2$) to assign them to each household i surveyed in the second round. First round predicted incomes for households surveyed in the second round are

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

where superscript U refers to the Upper bound estimate. The fraction of poor households in the first round who escape poverty can be expressed as $\Pr(\hat{y}_{i1}^{2U} < z \text{ and } y_{i2}^2 > z)$. Given that the method randomly draws from the empirical distribution of residuals, this procedure needs to be repeated R times and take an average of mobility estimates over the R replications.

Similarly, the lower bound estimate emerges from assuming perfect positive correlation between error terms. [Dang et al. \(2014\)](#) suggest using the estimates of the *scaled* residuals from the second round (represented by $\hat{\varepsilon}_{i2}^2$) to predict income in the first round

$$\hat{y}_{i1}^{2L} = \hat{\beta}_1' x_{i1}^2 + \frac{\hat{\sigma}_{\varepsilon_1}}{\hat{\sigma}_{\varepsilon_2}} \hat{\varepsilon}_{i2}^2 \quad (3)$$

where superscript L refers to the Lower bound estimate and where $\hat{\sigma}_{\varepsilon_1}$ and $\hat{\sigma}_{\varepsilon_2}$ are estimated standard errors for the two error terms ε_{i1} and ε_{i2} , respectively. The proportion of poor households who move out poverty can be obtained by computing the joint probability $\Pr(\hat{y}_{i1}^{2L} < z \text{ and } y_{i2}^2 > z)$. Given that we are not drawing randomly from any distribution, there is no need to repeat the procedure.

The technique was improved by [Dang and Lanjouw \(2013\)](#). The authors assume that the two error terms ε_{i1} and ε_{i2} have a bivariate normal distribution with a non-negative correlation coefficient ρ and standard deviations σ_{ε_1} and σ_{ε_2} , respectively. A point estimate of the probability of escaping poverty can be obtained as follow

$$P(y_{i1}^2 < z_1 \text{ and } y_{i2}^2 > z_2) = \Phi_2 \left(\frac{z_1 - \hat{\beta}_1' x_{i1}^2}{\hat{\sigma}_{\varepsilon_1}}, -\frac{z_2 - \hat{\beta}_2' x_{i1}^2}{\hat{\sigma}_{\varepsilon_2}}, -\rho \right) \quad (4)$$

The correlation ρ is unknown. Therefore, the authors use the age-cohort correlation of residuals obtained from cross-sectional data to produce an estimate of ρ .

In this paper, I propose an adaptation using the lower and upper bound estimates. Since the correlation term is likely to be non-negative, residuals in equations 2 and 3 are likely to “sandwich” the actual error term. Therefore, this paper suggests taking a weighted average of residuals to get point estimates of income mobility. First round predicted incomes under this method are

$$\hat{y}_{i1}^{2NP} = \hat{\beta}'_1 x_{i1}^2 + \left[\gamma \frac{\hat{\sigma}_{\varepsilon_1}}{\hat{\sigma}_{\varepsilon_2}} \varepsilon_{i2}^2 + (1 - \gamma) \varepsilon_{i1}^2 \right] \quad (5)$$

where superscript NP refers to the non-parametric estimate and $0 \leq \gamma \leq 1$. Non-parametric incomes are equal to lower bound estimates ($\hat{y}_{i1}^{2NP} = \hat{y}_{i1}^{2L}$) when $\gamma = 1$ and to upper bound ones ($\hat{y}_{i1}^{2NP} = \hat{y}_{i1}^{2U}$) when $\gamma = 0$. The proportion of poor households who escape poverty is given by the joint probability $Pr(\hat{y}_{i1}^{2NP} < z \text{ and } y_{i2}^2 > z)$. Once again, since we are randomly drawing from the empirical distribution of residuals from the first round, the procedure needs to be repeated R times.

This method requires an estimate of the unknown underlying weight γ . As noted, $\gamma = 0$ and $\gamma = 1$ correspond to the upper and lower bound mobility estimates, respectively. [Dang et al. \(2014\)](#) argue that the estimates would fall between these two extreme values. I can narrow down the gap between the lower and upper bound estimates by selecting $0 < \gamma < 1$. Therefore, I take “unweighted” averages of residuals (i.e., $\gamma = 0.5$) to validate estimates. By doing so, the paper is replacing the arbitrary assumption of normality and the need to estimate an unknown correlation term ρ by another arbitrary assumption of weighting lower and upper bound estimates equally (instead of, for example, setting γ equal to 0.4 or to 0.6). We could further refine this method by “weighting” residuals in equations 2 and 3 differently (i.e., $\gamma \neq 0.5$). As such, I also perform a sensitivity test in [Section 4](#) to analyze the robustness of results to changes in γ .

It is important to note that, by introducing an arbitrary assumption of balanced weights, the method also presents at least two advantages when compared to existing approaches. First, the adaptation introduced in this paper does not assume any given distribution for the disturbances. Second, unlike the parametric approach, this adaptation predicts the unobserved income in the first round of data rather than estimating probabilities of mobility in and out of poverty. This is also

crucial, since it allows to predict household-level income changes ($\widehat{\Delta y}_t^2 = y_{i2}^2 - \hat{y}_{i1}^{2NP}$) apart from the aggregated poverty dynamics obtained by the parametric approach.

3. Data and empirical approach for validation

I use the [SEDLAC](#) database for Peru as the primarily source information to validate income poverty dynamics. [SEDLAC](#) is a harmonized dataset of LAC's household surveys compiled by the World Bank and the Center for Distributive, Labor, and Social Studies (CEDLAS, for its acronym in Spanish) at the *Universidad Nacional de La Plata* in Argentina.⁴ The main objective of the [SEDLAC](#) dataset is to allow for cross-country comparisons of household incomes and a wide range of socioeconomic measures in LAC. The data cover more than 300 surveys in 18 countries in the region and represent more than 90 percent of the population in LAC.

To validate the non-parametric approach, I select a panel subsample of the [SEDLAC](#) data in Peru. This panel follows the same 1,075 households in a five-year period from 2007 to 2011, which allows us to validate the technique in periods of different lengths of time that include the same households. Validations are done for two periods. The first period covers one year from 2010 to 2011, while the second is a five-year period between 2007 and 2011. In this validation, the poor are defined as individuals with a harmonized per capita income lower than \$4 per person per day poverty line in 2005 \$PPP/day.

In all cases, mobility estimates are done by comparing the actual income in the second round and the predicted one—that arises from applying the method proposed in this paper using time invariant characteristics—in the first round. To avoid life cycle events that may invalidate the main time invariance assumption of the method, all estimates are based on data from households—as opposed to individuals—whose head is between 25 and 65 years of age (805 households). I apply 50 repetitions to the procedure and test how sensitive these results are to changes in the number of repetitions. I also choose $\gamma = 0.5$ and test the sensitivity of results to other values of γ . Following [Cruces et al. \(2015\)](#), I randomly divide all panel datasets into two subsamples and treat each subsample as a cross-section. Therefore, I estimate the coefficients in one of these subsamples in the first round (403 households) and apply those coefficients to the second subsample in the

⁴ [Bourguignon \(2015\)](#) and [Gasparini, Cicowiez, and Escudero \(2013\)](#) present a detailed description of the [SEDLAC](#) data.

second round (402 households). By treating each subsample as a cross-section, I avoid any bias that might arise from the use of panel datasets to validate the methodology.

I apply the following three specifications:

- Specification 1: includes household head age, age squared, gender, and years of education.
- Specification 2: adds regional fixed effects to the first specification.
- Specification 3: adds the interaction between variables in specification 1 and 2.

The paper also validates the methodology in other contexts by using the same panel datasets as the one presented by [Cruces et al. \(2015\)](#). These are three panel datasets for Chile, Nicaragua, and Peru: (i) the 1996 and 2006 National Socio-Economic Characterization Survey (CASEN in Spanish) panel survey from Chile; (ii) the 1998 and 2005 National Household Survey on Living Standards Measurement (EMNV in Spanish) panel survey from Nicaragua; and (iii) the 2008-2009 Peruvian National Household (ENAHO in Spanish) panel survey in Peru. These are nationally representative surveys that have information on education; employment; income in Chile and expenditure in Nicaragua and Peru; and housing, among others. The main objective of these surveys is to measure poverty and living conditions. For a detailed description of these datasets see [Cruces et al. \(2015\)](#). In these validations, the poor are defined as individuals with an official per capita expenditure (in Nicaragua and Peru) and income (in Chile) lower than the official poverty lines.

Unlike [Cruces et al. \(2015\)](#); [Dang et al. \(2014\)](#); and [Dang and Lanjouw \(2013\)](#), the harmonized [SEDLAC](#) data in Peru allow us to validate poverty dynamics using the same underlying income models and variables as the one frequently used in regional studies in LAC (for instance, [Ferreira et al. 2012](#) and [Vakis et al. 2016](#)). As such, I end the paper by presenting mobility results for LAC using the approach described above and the same underlying income models and variables as in the validation analysis. This allows us to implement the approach in most LAC countries for which data are available using the same underlying period and consistent concepts of income measures, datasets, and mobility measures across countries, thus considerably enlarging the universe of estimates and understanding of mobility trends in the region. I focus on intra-generational mobility in 17 LAC countries for which data are available between 2004 and 2015. In this analysis, the poor are individuals with a harmonized per capita income lower than \$4 per person and per day poverty line in 2005 \$PPP/day.

4. Validation results

4.1. Directional income mobility estimates

I start by comparing point estimates for the actual and simulated poverty headcount in round 1.⁵ Results are shown in [Table 1](#). Actual poverty estimates (and 95 percent confidence interval) in column 1 arise from using the panel dataset, while the predicted ones in columns 2, 3, and 4 result from applying the synthetic panel method. In general, the method performs well and actual point estimates are close to synthetic panel results.⁶

Models' predictive power increases substantially when moving from the first to the third specification (the R^2 doubles). However, unlike [Cruces et al. \(2015\)](#), point estimates are stable and they do not change drastically when increasing the sophistication level of the underlying income model.

I proceed to calculate the joint distribution of poverty transitions. [Tables 2](#) and [3](#) show the point estimate and 95 percent confidence interval for the actual panel mobility together with the point estimates from using synthetic panels in the two periods. Results are encouraging; with few exceptions, most of the point estimates are close to the actual estimates of poverty mobility and lie within the 95 percent confidence intervals. For instance, actual panel data in [Table 2](#) shows that about 17 to 25 percent remained in poverty in Peru between 2007 and 2011 (based on the 95 percent confidence interval), while according to the synthetic panel technique with interaction terms (specification 3) about 23 percent of the population were poor in both years. The two tables also suggest that the method performs well irrespective of the length of the panel.

Again, even when predictive power increases with the level of sophistication, results are stable to the selection of time invariant variables included in the underlying models. This is an important result for empirical applications at the regional level, as it suggests that the methodology can be implemented in a large set of countries simultaneously in contexts where there is a reduced set of comparable regressors available for all countries.

⁵ This is a useful exercise to perform in the absence of actual panel data since it allows us to check the validity of estimates from synthetic panels by comparing them with the actual poverty headcount that results from using cross sectional data in round 1.

⁶ Underlying income model estimations for Specification 1 are shown in the [appendix](#).

Table 4 shows conditional probabilities: the proportions of the population that move out of poverty or fall into poverty given their initial poverty status. The table presents results for the long panel data from 2007 to 2011 in panel A of the table and the short panel data from 2010 to 2011 in panel B of the table. Since both numerator and denominators are estimated in this figure, results are less accurate than those shown in Table 2 and 3 (Dang and Lanjouw 2013). Again, with few exceptions, most of the point estimates fall within the 95 percent confidence intervals. For instance, actual panel data in panel A shows that about four to 11 percent of the initial non-poor in 2007 were poor in 2011 (based on the 95 percent confidence interval), while according to the synthetic panel technique with interaction terms (specification 3) about 5 percent of the initial non-poor fell into poverty after five years.

4.2. Sensitivity checks

Given that we are randomly drawing from the empirical distribution of residuals from the first round, all previous estimates are based on 50 replications. Figure 1 analyzes the robustness of results to different number of repetitions for the 2007-2011 period.⁷ The graph shows results for specifications 3 increasing gradually the number of repetitions from 1 to 300. In general, estimates are robust to the number of repetitions; the figure suggests that estimations fall within 95 percent confidence intervals independently of the number of replications used. This result is extremely important because it simplifies the analysis by allowing us to obtain just only one predicted vector of incomes for the first round instead of 50 vectors as in Tables 2 and 3. Given this result, all remaining estimates are based on only one repetition.

All results have been based on an unweighted average (i.e., $\gamma = 0.5$) of residuals from the income models in the first and second round of data. Even when results have been encouraging and $\gamma = 0.5$ is a reasonable assumption based on empirical findings in the literature,⁸ one could think of different scenarios where this assumption does not hold. Dang and Lanjouw (2013) obtain a point estimate of poverty dynamics by assuming normality of the error terms and using the age-cohort correlation of residuals obtained from cross-sectional data in the first and second round to produce a point estimate of inter-generational poverty mobility. We could infer the value of γ by

⁷ Results are similar for the 2010-2011 period.

⁸ Residual correlations from panel data estimates are on average about 0.65 in Dang and Lanjouw (2013), about 0.50 in Dang et al. (2014), and about 0.45 in Cruces et al. (2015). Tables 2 and 3 show that correlation is about 0.50 in Peru between 2007 and 2011. These results provide support to the assumption of $\gamma = 0.5$.

following an approach similar to that suggested by [Dang and Lanjouw \(2013\)](#). However, this would move us away from the household-level analysis of the non-parametric approach to the age-cohort analysis performed under the parametric method. To avoid this, I test the sensitivity of the joint probabilities to changes in the value of the γ weights in [Figures 2 and 3](#). In general, the results are fairly stable to the value of the weights; except for extreme values, most estimates fall within the 95 percent confidence intervals.

[Figure 4](#) performs the same sensitivity analysis to the conditional probabilities shown in [Table 4](#). The method works well for $\gamma = 0.5$; most of the estimates fall within the 95 percent confidence interval when taking an unweighted average of residuals. However, since both numerator and denominators are estimated in this figure, results are less accurate than those shown in [Figures 2 and 3](#) ([Dang and Lanjouw 2013](#)) and extreme values fall outside the 95 percent confidence interval. Thus, caution should be exercised when selecting the weights to estimate conditional probabilities of moving in and out of poverty given the initial poverty status.

4.3. Subgroup and household-level poverty mobility

This section tests the performance of the methodology in measuring poverty dynamics for specific subgroups of the population. To do so, we estimate the same poverty transitions for several subgroups based on: region of residence, gender, age, and education of the household head. Subgroup synthetic panel results are based on parameters estimated from the entire sample. All estimations are based on specification 3 and one repetition is considered ($R = 1$).

[Figures 5 and 6](#) compare actual mobility from panel datasets with estimations based on the synthetic panel using the third specification in [Tables 2 and 3](#). All estimates are close to the 45-degree line, suggesting that the technique works well for estimating poverty dynamics profiles based on subgroups of the overall population independently of the panel length.

[Figure 7](#) validates the methodology for estimating non-anonymous growth incidence curves (GIC). These curves show household-level welfare growth by deciles of the first-round welfare distribution.⁹ All estimations are based on specification 3 and are compared with the actual non-anonymous GIC from panel data. All estimated GICs from synthetic panels are close to the actual income growth. These results are extremely relevant since, unlike the parametric approach

⁹ As opposed to anonymous GICs which show decile-level (or any other percentile) welfare growth by deciles (or any other percentile) of the first-round welfare distribution.

developed by [Dang and Lanjouw \(2013\)](#), results in these figures suggest that the method proposed in this paper performs well in predicting household welfare changes (i.e., $\widehat{\Delta y}_t^2 = y_{i2}^2 - \hat{y}_{i1}^{2NP}$) apart from predicting joint probabilities of moving in and out of poverty.

[Figures 5](#) through [7](#) show that the method works well for predicting income mobility for several subgroups of the population. However, it is crucial to understand whether “household-level” transitions predicted from the synthetic panel methodology are similar to transitions observed in the actual panel data.¹⁰ [Table 5](#) presents the proportion of households based on the actual mobility from true panels and the proportion of households based on predicted mobility based on synthetic panels. Results indicate that methodology works well for predicting household level mobility; most of the estimates are in the main diagonal. However, caution should be taken when predicting “household-level” mobility, since about 20 percent of the individuals are not classified correctly.

4.4. Validation in a different context

This section validates in other contexts by using the same panel datasets for Chile, Nicaragua, and Peru as the one introduced by [Cruces et al. \(2015\)](#). [Figure 8](#) shows the conditional probability of poverty transition for the three countries and for different values of γ when implementing their third specification.¹¹ The figure presents the point estimate and 95 percent confidence intervals for the actual panel mobility together with the point estimates from using synthetic panels in the three countries.

With few exceptions, most of the point estimates using $\gamma = 0.5$ are close to the actual estimates of poverty mobility and lie within the 95 percent confidence intervals. For instance, actual panel data shows that between 5 and 12 percent of the initial non-poor in 2001 in Nicaragua fell into poverty in 2005 (based on the 95 percent confidence interval), while according to the synthetic panel technique about 9 percent of the original non-poor population fell into poverty between both years. Chile performs the least well. The conditional upward mobility estimate lies about 8 percentage points outside the 95 percent confidence interval of the actual upward mobility in this country—though, the estimation of the conditional probability of falling into poverty lies

¹⁰ This validation has not been done by [Cruces et al. \(2015\)](#) and [Dang et al. \(2013, 2014\)](#).

¹¹ Specification 3 in [Cruces et al. \(2015\)](#) includes: (i) household head age, age squared, gender, years of education, and ethnicity; (ii) geographical controls and regional fixed effects; and (iii) the interaction between variables in (i) and (ii).

within the confidence interval. This might be explained by the relatively longer period between rounds and smaller sample size. However, the difference is only 11 percent in relative terms for the conditional upward mobility. Thus, I think that these results confirm that the methodology also works well in other contexts.

5. An application for LAC

This section presents mobility results for the whole LAC region using the approach described in [Section 2](#). To this end, I use the [SEDLAC](#) database of LAC household surveys and I focus on intra-generational mobility measures for the whole region, as well as for 17 LAC countries for which data are available between 2004 and 2015. These years coincide with a period of sustained economic growth and marked poverty reduction ([Cord et al. 2017](#), [Vakis et al. 2016](#)).

[Figure 9](#) shows two conditional types of transitions: (i) the proportion of poor who escape poverty (defined as those with income lower than \$4 PPP per day); and (ii) the proportion of non-poor who fall into poverty. As the figure suggests, LAC has experienced a dramatic mobility in the last 11 years. There is considerably more upward than downward mobility: 43 percent of the initial poor escaped poverty by 2015, while about 10 percent of those whose initial incomes were higher than the poverty line fell into poverty during the last 11 years.

These trends vary across countries. For instance, less than a quarter of the original poor escaped poverty in Guatemala and Honduras during the last 11 years, which compares with a significantly higher conditional upward mobility of more than 70 percent in Uruguay and Chile. More important, despite the strong poverty reduction in the last decade in LAC, downward income mobility is particularly worrying in some countries in the region. For instance, more than 10 percent of the initial non-poor fell into poverty in Nicaragua, Mexico, Honduras, and Guatemala.

The key point of this section is not to provide an extensive analysis of mobility in the region but rather to showcase the type of new insights that a researcher can obtain by applying the technique described in the paper. Indeed, cross sectional data using the same years and data sources would not have allowed an understanding of the welfare dynamics that the technique allows disentangling as above. Using cross sectional data alone would only reveal that between 2004 and 2015, poverty in the region went down from about 40 to 24 percent; but we would not know the nature of those movements.

6. Conclusion

Due to an increasing interest in measuring intra-generational mobility, and since panel datasets are rarely available or have many limitations, there has been a growing literature that studies mobility using cross-sectional data that do not follow individuals over time. Synthetic panel is the most recent technique and allows estimating the unobserved welfare level in the first round of cross sectional data for all households interviewed in the second round. A non-parametric version of the technique yields lower and upper bound mobility estimates depending on the assumptions made about the correlation between the first and second round error terms in the underlying regressions. On the other hand, a parametric version of the method produces a point estimate of poverty mobility.

By taking a weighted average of the estimated residuals in the underlying models, this paper proposes a non-parametric adaptation of the parametric point estimate of mobility. Unlike the parametric technique, this adaptation does not require any assumption regarding the distribution of the error terms in the underlying welfare models. The method is validated by producing welfare mobility estimates out of cross-sectional data and by comparing them with actual mobility from panel datasets in three LAC countries: Chile, Nicaragua, and Peru. Results suggest that the method performs well; most estimates are close to the actual panel mobility. Moreover, these results confirm that the method yields reliable estimates of household welfare growth, apart from the joint probabilities of poverty dynamics. Finally, the paper also suggests that the methodology can be easily applied to regional analysis, since it does not require a large number of explanatory variables in the underlying welfare model estimations.

References

- Bourguignon, F. 2015. "Appraising income inequality databases in Latin America" *The Journal of Economic Inequality* 13 (4): 557–578.
- CEDLAS, and World Bank. 2015. "SEDLAC: Socio-Economic Database for Latin America and the Caribbean." *SEDLAC*. August. <http://sedlac.econo.unlp.edu.ar/eng/>.
- Cord, Louise, Oscar Barriga-Cabanillas, Leonardo Lucchetti, Carlos Rodríguez-Castelán, Liliana D. Sousa, and Daniel Valderrama. 2017. "Inequality Stagnation in Latin America in the Aftermath of the Global Financial Crisis." *Review of Development Economics* 21 (1): 157–81. doi:10.1111/rode.12260.
- Cruces, Guillermo, Peter Lanjouw, Leonardo Lucchetti, Elizaveta Perova, Renos Vakis, and Mariana Viollaz. 2015. "Intra-Generational Mobility and Repeated Cross-Sections: A Three-Country Validation Exercise." *Journal of Economic Inequality* 13 (2): 161–79.
- Dang, Hai-Anh, Peter Lanjouw, Jill Luoto, and David McKenzie. 2014. "Using Repeated Cross-Sections to Explore Movements into and out of Poverty". *Journal of Development Economics*. 107, 112–128.
- Dang, Hai-Anh and Peter Lanjouw. 2013. "Measuring Poverty Dynamics with Synthetic Panels Based on Cross-Sections." World Bank Policy Research Working Paper 6540.
- Ferreira, Francisco H. G., Julian Messina, Jamele Rigolini, Luis-Felipe López-Calva, Maria Ana Lugo, and Renos Vakis. 2013. *Economic Mobility and the Rise of the Latin American Middle Class*. Washington, DC: World Bank.
- Gasparini, Leonardo, Martín Cicowiez, and Walter Sosa Escudero. 2013. *Pobreza y desigualdad en América Latina: conceptos, herramientas y aplicaciones*. La Plata, Argentina: Temas Grupo Editorial Srl.
- Vakis, Renos; Jamele Rigolini; and Leonardo Lucchetti. 2016. *Left behind: chronic poverty in Latin America and the Caribbean*. Washington, DC; World Bank Group.

Tables and figures

Table 1: Actual and Simulated Poverty Headcount in the First Round

Status in Round 1	Actual	Synthetic Panel		
	[1]	[2]	[3]	[4]
<i>Panel A: Peru 2007</i>				
Poverty Rate	41.5 (40.5, 42.5)	37.6	38.8	39.1
<i>Panel A: Peru 2010</i>				
Poverty Rate	29.9 (28.9, 30.7)	24.9	28.5	29.9
R-squared Panel A	.	0.24	0.45	0.49
R-squared Panel B	.	0.22	0.40	0.43
Observations	403	403	403	403

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: R-squared is calculated for opposite halves of the total first round sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results in column [1] show actual panel poverty. Column [2] shows a simple model with household time invariant characteristics. Column [3] adds region fixed effects. Column [4] adds interactions between household time invariant characteristics and fixed effects. Estimations are based on 50 repetitions and $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4 a day. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Table 2: Transition Matrices – Repeated Cross Sections vs. Panel Data
Peru 2007-2011

Status in 2007, 2011	Actual	Synthetic Panel		
	[1]	[2]	[3]	[4]
Poor, Poor	21.1 (17.1, 25.1)	22.0	23.4	22.8
Poor, Non-poor	20.4 (16.4, 24.3)	15.7	15.4	16.4
Non-poor, Poor	4.5 (2.4, 6.5)	3.7	2.2	2.8
Non-poor, Non-poor	54.0 (49.1, 58.8)	58.7	59.0	58.0
R-squared	.	0.24	0.45	0.49
Residual Correlation	.	0.57	0.48	0.48
Observations	403	403	403	403

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: R-squared is calculated for opposite halves of the total 2007 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results in column [1] show actual panel mobility. Column [2] shows a simple model with household time invariant characteristics. Column [3] adds region fixed effects. Column [4] adds interactions between household time invariant characteristics and fixed effects. 95 percent confidence interval between parentheses. Estimations are based on 50 repetitions and $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Table 3: Transition Matrices – Repeated Cross Sections vs. Panel Data
Peru 2010-2011

Status in 2010, 2011	Actual	Synthetic Panel		
	[1]	[2]	[3]	[4]
Poor, Poor	18.4 (14.6, 22.2)	17.4	20.4	20.5
Poor, Non-poor	11.4 (8.3, 14.5)	7.4	8.1	9.4
Non-poor, Poor	7.2 (4.6, 9.7)	8.2	5.2	5.1
Non-poor, Non-poor	62.9 (58.2, 67.6)	67.0	66.3	65.0
R-squared	.	0.22	0.40	0.43
Residual Correlation	.	0.63	0.55	0.54
Observations	403	403	403	403

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: R-squared is calculated for opposite halves of the total 2010 sample. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Results in column [1] show actual panel mobility. Column [2] shows a simple model with household time invariant characteristics. Column [3] adds region fixed effects. Column [4] adds interactions between household time invariant characteristics and fixed effects. 95 percent confidence interval between parentheses. Estimations are based on 50 repetitions and $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Table 4: Transition Matrices – Repeated Cross Sections vs. Panel Data
Conditional Probability

Conditional Mobility	Actual	Synthetic Panel		
	[1]	[2]	[3]	[4]
<i>Panel A: Peru 2007 – 2011</i>				
Proportion of poor in 2007 who moved out of poverty in 2011	49.1 (41.4, 56.7)	41.7	39.7	41.8
Proportion of non-poor in 2007 who moved into poverty in 2011	7.7 (4.2, 11.0)	5.9	3.7	4.7
<i>Panel B: Peru 2010 – 2011</i>				
Proportion of poor in 2010 who moved out of poverty in 2011	38.3 (29.5, 47.0)	29.9	28.4	31.3
Proportion of non-poor in 2010 who moved into poverty in 2011	10.3 (6.7, 13.8)	10.9	7.3	7.2
Observations	403	403	403	403

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: R-squared is calculated for opposite halves of the total 2007 sample in Panel A and 2010 in Panel B. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Column [1] shows actual panel mobility. Column [2] shows a simple model with household time invariant characteristics. Column [3] adds region fixed effects. Column [4] adds interactions between household time invariant characteristics and fixed effects. 95 percent confidence interval between parentheses. Estimations are based on 50 repetitions and $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Table 5: Household-level Dynamics and Profiling– Repeated Cross Sections vs. Panel Data
Peru 2007-2011 and 2010-2011

Actual Mobility	Synthetic Panels				Diagonal
	Poor, Poor	Poor, Non-poor	Non-poor, Poor	Non-poor, Non-poor	
<i>Panel A: Peru 2007-2011</i>					
Poor, Poor	19.4	0.0	1.7	0.0	79
Poor, Non-poor	0.0	10.7	0.0	9.7	
Non-poor, Poor	3.5	0.0	1.0	0.0	
Non-poor, Non-poor	0.0	5.7	0.0	48.3	
<i>Panel B: Peru 2010-2011</i>					
Poor, Poor	15.7	0.0	2.7	0.0	78
Poor, Non-poor	0.0	3.5	0.0	8.0	
Non-poor, Poor	5.2	0.0	2.0	0.0	
Non-poor, Non-poor	0.0	6.0	0.0	57.0	

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are based on the third specification. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. No repetitions are done ($R = 1$). Estimations are based on $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Figure 1. Poverty Dynamics using Different Number of Replications
Peru 2007 and 2011

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. The horizontal axis shows the number of repetitions. The gray area represents the 95-confidence interval for the actual estimations. Dashed lines show the model with interactions between household time invariant characteristics and regional fixed effects. Estimations are based on $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Figure 2. Poverty Dynamics using Different Weights γ
Peru 2007-2011

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. The horizontal axis shows the weights. The gray area represents the 95-confidence interval for the actual estimations. Dashed lines show the model with interactions between household time invariant characteristics and regional fixed effects. Solid lines represent actual mobility from panel data. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. No repetitions are done ($R = 1$).

Figure 3. Poverty Dynamics using Different Weights γ
Peru 2010-2011

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. The horizontal axis shows the weights. The gray area represents the 95-confidence interval for the actual estimations. Dashed lines show the model with interactions between household time invariant characteristics and regional fixed effects. Solid lines represent actual mobility from panel data. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. No repetitions are done ($R = 1$).

Figure 4. Conditional Poverty Dynamics using Different Weights γ

Panel A: Peru 2007-2011

Proportion of Poor in 2007 who Escaped Poverty in 2011

Proportion of Non-poor in 2007 who Fall into poverty in 2011

Panel B: Peru 2010-2011

Proportion of Poor in 2010 who Escaped Poverty in 2011

Proportion of Non-poor in 2010 who Fall into poverty in 2011

Data source: INEI - Peruvian National Household Survey (ENAHU), 2007, 2010, and 2011. Note: R-squared is calculated for opposite halves of the total 2007 sample in Panel A and 2010 in Panel B. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. The gray area represents the 95-confidence interval for the actual estimations. Solid lines represent actual mobility from panel data. Results are based on the third specification. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. No repetitions are done ($R = 1$).

Figure 5. Poverty Dynamics by Subgroups of the Population
Peru 2007 and 2011

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. No repetitions are done ($R = 1$). Estimations are based on $\gamma = 0.5$. The 45-degree line shows actual panel mobility. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Figure 6. Poverty Dynamics by Subgroups of the Population
Peru 2010 and 2011

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. No repetitions are done ($R = 1$). Estimations are based on $\gamma = 0.5$. The 45-degree line shows actual panel mobility. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Figure 7. Non-anonymous Growth Incidence Curves (GIC)
(a) Peru 2007 and 2011

(b) Peru 2010 and 2011

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Incomes are expressed in 2005 \$PPP/day. No repetitions are done ($R = 1$). Estimations are based on $\gamma = 0.5$.

Figure 8. Conditional Poverty Dynamics using Different Weights γ

Panel A: Peru 2008-2009

Proportion of Poor in 2008 who Escaped Poverty
in 2009

Proportion of Non-poor in 2008 who Fall into
poverty in 2009

Panel B: Nicaragua 2001-2005

Proportion of Poor in 2001 who Escaped Poverty
in 2005

Proportion of Non-poor in 2001 who Fall into
poverty in 2005

Panel C: Chile 1996-2006

Proportion of Poor in 1996 who Escaped Poverty
in 2006

Proportion of Non-poor in 1996 who Fall into
Poverty in 2006

Data source: EMNV, 2001-2005 in Panel A and CASEN, 1996-2006 in Panel B. Note: R-squared is calculated for opposite halves of the total 2001 sample in Panel A and 1996 in Panel B. Results are constrained to the panel sample of households whose heads are between 25 and 55 years old. The gray area represents the 95-confidence interval for the actual estimations. Solid lines represent actual mobility from panel data. Results are based on the third specification in [Cruces et al. \(2015\)](#). Poor are those individuals with a per capita expenditure and income lower than the official poverty lines. No repetitions are done ($R = 1$).

Figure 9: Conditional Intra-generational Mobility by Country in LAC, 2004-2015
 Panel A: Proportion of Poor in 2004 who Escaped Poverty in 2015

Panel B: Proportion of Non-poor in 2004 who fall into Poverty in 2015

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to households whose heads are between 25 and 65 years old. The underlying models include interactions between household time invariant characteristics and regional fixed effects. No repetitions are done. Estimations are based on $\gamma = 0.5$. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Appendix

Table A: Log Income Estimates Based on Specification 1
Peru 2007, 2010, and 2011

Dependent Variable: Log Per Capita Income	Years		
	2007	2010	2011
	[1]	[2]	[3]
Male	-0.177*	-0.161*	-0.263***
	[0.107]	[0.096]	[0.085]
Years of Education	0.099***	0.079***	0.085***
	[0.009]	[0.008]	[0.008]
Age	-0.045	0.017	0.031
	[0.038]	[0.040]	[0.038]
Age Squared	0.001*	0.000	0.000
	[0.000]	[0.000]	[0.000]
Constant	4.616***	3.655***	3.453***
	[0.818]	[0.906]	[0.860]
Number of households	403	403	402

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old.