



Poverty & Equity Global Practice Working Paper 163

# WHAT CAN WE (MACHINE) LEARN ABOUT WELFARE DYNAMICS FROM CROSS-SECTIONAL DATA?

Leonardo Lucchetti  
August 2018

## ABSTRACT

This paper implements a machine learning approach to estimate intra-generational economic mobility using cross-sectional data. A Least Absolute Shrinkage and Selection Operator (Lasso) procedure is applied to explore poverty dynamics and household-level welfare growth in the absence of panel data sets that follow individuals over time. The method is validated by sampling repeated cross-sections of actual panel data from Peru. In general, the approach performs well at estimating intra-generational poverty transitions; most of the mobility estimates fall within the 95 percent confidence intervals of poverty mobility from the actual panel data. The validation also confirms that the Lasso regularization procedure performs well at estimating household-level welfare growth between two years. Overall, the results are sufficiently encouraging to estimate economic mobility in settings where panel data are not available or, if they are, to improve panel data when they suffer from serious non-random attrition problems.

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 contribute to development policy discussions around the world. The authors may be contacted at [fadoho@worldbank.org](mailto:fadoho@worldbank.org) and [llucchetti@worldbank.org](mailto:llucchetti@worldbank.org).

The Poverty & Equity Global Practice 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.

– Poverty & Equity Global Practice Knowledge Management & Learning Team

# What Can We (Machine) Learn about Welfare Dynamics from Cross-Sectional Data?

Leonardo Lucchetti  
The World Bank

*Keywords:* Poverty; Poverty transitions; LASSO; Machine learning, Synthetic panels; Welfare dynamics.

*JEL classification:* O15, I32.

*Sector Board:* POV

---

Leonardo Lucchetti ([llucchetti@worldbank.org](mailto:llucchetti@worldbank.org)) is Senior Economist with the Poverty Global Practice, World Bank. I am grateful to Tanida Arayavechkit, Monserrat Bustelo, Oscar Calvo, Andres Castaneda, Jonathan Hersh, Daniel Lederman, David Newhouse, Ana Maria Oviedo, Alberto Rodriguez, Joana Silva, Emmanuel Skoufias, Liliana Sousa, and participants at World Bank seminars whose suggestions greatly improved earlier drafts of the paper. All remaining errors are mine.

## 1. Introduction

There has been a considerable increase in the number of countries that have developed the necessary tools to measure poverty in recent years. In addition, a large body of research has proposed standardized methods to compare poverty across countries, as well as to monitor poverty evolution at a regional and global level (Ravallion, Datt, and van de Walle 1991; Chen and Ravallion 2001; Ravallion, Chen, and Sangraula 2009; Jolliffe and Prydz 2016; Ferreira et al. 2016; Castaneda et al., forthcoming). The rapid expansion of household surveys at frequent intervals and comparable over time and across countries has facilitated poverty monitoring in the developing world; coverage increased from 13 countries in the 1990s to over 60 countries in 2011 (Serajuddin et al. 2015). However, most of the micro data available are cross-sectional that do not track individuals and households over time and therefore only provide aggregate poverty trends. Panel datasets that follow individuals over several periods of time are rarely available, which limits the understanding of the underlying factors behind movements out of poverty, the dynamics into poverty, and the duration of poverty experienced by a group of individuals.

This paper introduces a supervised machine learning method to estimate intra-generational economic mobility using cross-sectional data.<sup>1</sup> The method estimates parameters in the first round of cross sectional data by means of the Lasso regularization process (Tibshirani 1996). A cross-validation method is used to evaluate the out-of-sample predictive performance of the model in the first round of data. These estimated parameters are then used to predict a point estimate of the unobserved income in the first round for all households surveyed in the second round and estimate intra-generational poverty transitions in the absence of panel data. This approach is validated by comparing estimates from cross-sectional data with those from actual panel data from Peru.

A large body of research on the subject has emerged in recent years. “Synthetic Panels”, developed by Dang et al. (2014), is the most recent one.<sup>2</sup> The authors estimated a (log) income<sup>3</sup> model in both the first and second rounds of cross-sectional data, including time-invariant

---

<sup>1</sup> Mullainathan and Jann Spiess (2017) present a detailed description of the use of machine learning methods in economics. Supervised machine learning consists in producing good *predictions* of a variable  $y$  from the values of  $x$ , as opposed to the classical econometric problem of obtaining good estimates of parameters  $\beta$  that describe the relation between both variables. *Supervised* machine learning refers to those situations where a value of  $y$  is observed for each value of  $x$ . Conversely, we do not observe a value of  $y$  for each value of  $x$  under the *unsupervised* machine learning.

<sup>2</sup> The Synthetic panel method builds on the poverty mapping technique developed by Elbers, Lanjouw, and Lanjouw (2003).

<sup>3</sup> For simplicity, I will refer to income as the welfare measure in this paper.

covariates and retrospective regressors. Parameters estimated in the first round are then used to predict the unobserved income in the first round for all households interviewed in the second round. Depending on the assumptions introduced with respect to the correlation between the error terms in the underlying regressions in both rounds, this “non-parametric” approach generates an upper and lower bound poverty mobility using cross sectional data. The methodology was validated in Chile, Nicaragua, and Peru by [Cruces et al. \(2015\)](#), while [Ferreira et al. \(2012\)](#) predicted intra-generational poverty mobility in 18 countries in Latin America and the Caribbean (LAC) by implementing the lower bound estimates with harmonized cross-sectional micro data.

By assuming normality of the error terms in the underlying regressions and by using the age-cohort correlation of residuals from cross-sections, [Dang and Lanjouw \(2013\)](#) produced a point estimate of intra-generational poverty mobility—as opposed to upper and lower bound estimates. This “parametric” method was validated by the authors using panel data from five countries. The method was applied by [Dang and Lanjouw \(forthcoming\)](#) to study poverty dynamics in India, by [Dang and Dabalén \(forthcoming\)](#) to analyze whether growth has been pro-poor in 21 countries in Africa, and by [Vakis, Rigolini, and Lucchetti \(2016\)](#) to analyze chronic poverty in 17 LAC countries for which harmonized cross-sectional micro data exist.

[Lucchetti \(2017\)](#) developed a “non-parametric” point estimate of the unobserved household income in the first round for all households surveyed in the second round of cross-sectional data. To this end, the author calculates a weighted average of the residuals obtained in the upper and lower bound estimates. This approach is validated using actual panel data from Chile, Nicaragua, and Peru, and applied in 17 LAC countries for which harmonized micro data are available. This non-parametric point estimate requires an unknown underlying weight  $\gamma$  when computing the weighted average of lower and upper bound residuals. The author introduces an ad-hoc assumption by weighting lower and upper bound estimates equally—i.e., setting  $\gamma=0.5$ —and performs a sensitivity test of results to changes in the value of  $\gamma$ .

The machine learning approach introduced in this paper presents several strengths and uses less restrictive assumptions than similar studies previously developed. First, the method does not use estimated residuals from regressions. Therefore, no normal distribution of error terms in the underlying income regressions needs to be assumed.<sup>4</sup> Second, this approach does not introduce any arbitrary underlying weight  $\gamma$  as in [Lucchetti \(2017\)](#) and it does not require the estimation of

---

<sup>4</sup> The assumption of normality of error terms is rejected in Vietnam and Indonesia by [Dang et al. \(2014\)](#).

the age-cohort correlation of residuals from cross-sections as in [Dang and Lanjouw \(2013\)](#). Third, unlike [Dang et al. \(2014\)](#), this machine learning approach also predicts point estimates of income mobility—as opposed to just predicting probabilities of poverty transitions.

This paper contributes to the growing empirical literature on the use of machine learning to predict economic well-being. [Engstrom et al. \(2017\)](#) use regularization processes together with satellite images to estimate poverty at a high level of geographical disaggregation in Sri Lanka. [Babenko et al. \(2017\)](#) train Convolutional Neural Networks and use satellite images to also estimate the spatial distribution of poverty in Mexico. [Afzal et al. \(2015\)](#) test the accuracy of poverty estimations using machine learning methods—also combined with satellite data—in Pakistan and Sri Lanka. Finally, [McBride and Nichols \(2016\)](#) focus on machine learning techniques to improve targeting tools to identify potential program beneficiaries.

Results in this paper reveal that the Lasso regularization process performs well at predicting intra-generational poverty transitions in the context of the Peruvian data. Most of the estimates fall within the 95 percent confidence intervals of the joint and conditional probability of poverty mobility of the true panel data. The paper also finds that the method does well at predicting household-level income growth—and not just poverty transitions—between the two rounds of cross-sectional data. The analysis reveals that these predictions can be further improved by randomly drawing observed incomes from the distribution in round 1 and allocating them to each household surveyed in round 2 based on their position in the distribution of predicted income that results from the Lasso regularization approach described in this paper.

The next [section](#) summarizes all the Synthetic panel approaches, as well as the machine learning method proposed in this paper. [Section 3](#) presents the main data used. [Section 4](#) discusses the validation results. Finally, [Section 5](#) concludes.

## 2. Methodology<sup>5</sup>

### 2.1. Non-parametric Synthetic panels

Assume two rounds of cross-sectional data. We call  $y_{it}$  household's  $i$  log per capita income in moment  $t$ ,  $x_{it}$  a vector of household characteristics for household  $i$  in round  $t$ , and  $z$  the poverty line. Characteristics included in  $x_{it}$  are variables whose first round value can be inferred for all

---

<sup>5</sup> This section largely relies on [Dang and Lanjouw \(2013\)](#), [Dang et al. \(2014\)](#), [Cruces et al. \(2015\)](#), [Vakis et al. \(2016\)](#), and [Lucchetti \(2017\)](#).

households surveyed in the second round of data. These characteristics include: (i) time-invariant variables such as gender of the head of the household if his/her identity remains constant between the rounds of data; (ii) deterministic variables such as age; and (iii) retrospective variables such as whether a household surveyed in the second round had an asset in the first round (Cruces et al. 2015, Dang and Lanjouw 2018). The relationship between income and a set of time invariant characteristics can be expressed as

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

where  $\varepsilon_{it}$  is an error term and  $x_{it}$  is a vector of  $K$  regressors whose first element is equal to one so that the first element of  $\beta_t$  is the intercept of the model.

We introduce superscripts to refer to observations surveyed in each moment in time. As such, the objective is to estimate, for a household  $i$  interviewed in round 2, the change of incomes between the two rounds of data:  $\Delta y_i^2 = y_{i2}^2 - y_{i1}^2$ , where  $y_{i1}^2$  and  $y_{i2}^2$  are the first and second round incomes of household  $i$  surveyed in round 2, respectively. Similarly, we can also estimate all poverty dynamics: the joint probability of a household  $i$  surveyed in round 2 of escaping poverty in round 2 ( $\Pr(y_{i1}^2 < z \text{ and } y_{i2}^2 > z)$ ), remaining poor ( $\Pr(y_{i1}^2 < z \text{ and } y_{i2}^2 < z)$ ), becoming poor ( $\Pr(y_{i1}^2 > z \text{ and } y_{i2}^2 < z)$ ), and remaining non-poor ( $\Pr(y_{i1}^2 > z \text{ and } y_{i2}^2 > z)$ ).<sup>6</sup>

This can be easily done with panel data, since all households are interviewed in both rounds (i.e.,  $y_{i1}^2$  is known for every household  $i$  interviewed in round 2). However, these datasets are rarely available and costly to collect. Alternatively, Synthetic panels allow predicting the first round “unobserved” incomes of households surveyed in the second round by multiplying their time-invariant characteristics and the first-round Ordinary Least Squares (OLS) estimates of parameters  $\hat{\beta}_1^{OLS}$  that solve the optimization problem

$$\hat{\beta}_1^{OLS} = \underset{\beta_1}{\operatorname{argmin}} [\sum_{i=1}^{N_1} (y_{i1}^1 - \beta_1' x_{i1}^1)^2] = \underset{\beta_1}{\operatorname{argmin}} [RSS] \quad (2)$$

where  $y_{i1}^1$  is the first-round log income of household  $i$  surveyed in round 1,  $N_1$  indexes the number of observations in round 1, and RSS refers to the residual sum of squares. The three non-parametric approaches differ in the treatment given to the correlation between the error terms in the first and second rounds of cross-sectional data, which is likely to be non-negative according to Dang et al. (2014).

---

<sup>6</sup> For simplicity, I will only focus on the probability of escaping poverty in this section.

Upper bound estimates assume no correlation between the first and second round error terms. The authors propose to estimate first round incomes of those households interviewed in the second round of data by drawing randomly with replacement from the empirical distribution of first round estimated residuals (denoted as  $\tilde{\varepsilon}_{i1}^2$ ). In this case, the upper bound prediction of the first-round incomes for households surveyed in the second round is

$$\hat{y}_{i1}^{2U} = \hat{y}_{i1}^2 + \tilde{\varepsilon}_{i1}^2 \quad (3)$$

where  $\hat{y}_{i1}^2$  is the product between time-invariant characteristics and the first-round OLS estimates of parameters:  $\hat{y}_{i1}^2 = \hat{\beta}_1^{OLS'} x_{i1}^2$ . Once incomes are predicted, we can then calculate the joint probability of a household  $i$  surveyed in round 2 of being poor in round 1 and escape poverty in round 2,  $\Pr(\hat{y}_{i1}^{2U} < z \text{ and } y_{i2}^2 > z)$ , as well as the income change between both periods  $\Delta y_i^{2U} = y_{i2}^2 - \hat{y}_{i1}^{2U}$ . Since predictions arise from a random draw of the empirical distribution of residuals, the method needs to be repeated  $R$  times and results averaged over these  $R$  replications.<sup>7</sup>

Lower bound estimates on the other hand assume perfect positive correlation between the first and second round error terms. The authors propose to estimate first round incomes of those households interviewed in the second round of data by using the estimates of the scaled residuals from the second-round regression (denoted as  $\hat{\varepsilon}_{i2}^2$ ). The lower bound predictions are

$$\hat{y}_{i1}^{2L} = \hat{y}_{i1}^2 + \frac{\hat{\sigma}_{\varepsilon_1}}{\hat{\sigma}_{\varepsilon_2}} \hat{\varepsilon}_{i2}^2 \quad (4)$$

where  $\hat{\sigma}_{\varepsilon_1}$  and  $\hat{\sigma}_{\varepsilon_2}$  are estimated standard errors for the two error terms  $\varepsilon_{i1}$  and  $\varepsilon_{i2}$ , respectively. The joint probability of a household  $i$  surveyed in round 2 of being poor in round 1 and escape poverty in round 2 is given by  $\Pr(\hat{y}_{i1}^{2L} < z \text{ and } y_{i2}^2 > z)$ , while the change in incomes between both periods is  $\Delta y_i^{2L} = y_{i2}^2 - \hat{y}_{i1}^{2L}$ . Since the method is not randomly drawing from any the empirical distribution of residuals, there is no need to repeat the procedure  $R$  times.

The third non-parametric point estimate proposed by [Lucchetti \(2017\)](#) is an adaptation of the lower and upper bound estimations. The author suggests computing a weighted average of the residuals to get a point estimate of mobility. First round non-parametric predicted incomes are

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

---

<sup>7</sup> [Cruces et al. \(2015\)](#) show that results are robust to the number of repetitions  $R$ .

where  $0 \leq \gamma \leq 1$ . The joint probability of a household  $i$  surveyed in round 2 of being poor in round 1 and escape poverty in round 2 is given by  $\Pr(\hat{y}_{i1}^{2NP} < z \text{ and } y_{i2}^2 > z)$ , while the change in incomes between both periods is  $\Delta y_i^{2NP} = y_{i2}^2 - \hat{y}_{i1}^{2NP}$ . Since upper bound residuals are used, the method needs to be repeated  $R$  times.<sup>8</sup> The lower bound estimates can be obtained by setting  $\gamma = 1$ , while the upper bound estimates emerge from setting  $\gamma = 0$ . Based on residual correlations estimated from panel data in the literature, the author sets  $\gamma = 0.5$  and test the sensitivity of results to changes in the value of the  $\gamma$ .

## 2.2. A parametric Synthetic panel

[Dang and Lanjouw \(2013\)](#) propose a parametric point estimate of the intra-generational poverty mobility. The authors assume a bivariate normal distribution for the error terms with a non-negative correlation coefficient  $\rho$ . Thus, a point estimate of the probability of moving out of poverty is

$$\Pr(y_{i1}^2 < z \text{ and } y_{i2}^2 > z) = \Phi \left( \frac{z - \hat{\beta}_1^{OLS'} x_{i1}^2}{\hat{\sigma}_{\varepsilon_1}}, \frac{z - \hat{\beta}_2^{OLS'} x_{i1}^2}{\hat{\sigma}_{\varepsilon_2}}, -\rho \right) \quad (6)$$

where  $\hat{\beta}_2^{OLS}$  are the second-round OLS parameter estimates. A parametric lower bound estimate can be obtained by setting  $\rho = 1$ , while the upper bound estimate emerges from setting  $\rho = 0$ . The authors suggest estimating an age-cohort correlation of residuals using cross-sectional data to obtain an estimation of the unknown parameter  $\rho$ .

## 2.3. A Machine Learning approach based on the Lasso regularization method

This paper applies a Lasso regularization method to estimate intra-generation poverty mobility and household-level income growth using cross-sectional data. The Lasso procedure is one of the most popular machine learning methods among economists and consists on minimizing a quadratic loss function plus the sum of the absolute value of the coefficients ([Mullainathan and Jann Spiess 2017](#)). The paper proposes to estimate parameters in the first round of cross-sectional data by solving the optimization problem

$$\hat{\beta}_{1\lambda}^{LASSO} = \underset{\beta_1}{\operatorname{argmin}} \left[ RSS + \lambda \sum_{s=1}^K |\beta_{s1}| \right] \quad (7)$$

---

<sup>8</sup> The author shows that results are robust to the number of repetitions  $R$ .

The estimation depends on the value of the “shrinkage” factor  $\lambda$ . Whenever  $\lambda \rightarrow 0$ , the objective function will become the OLS objective function in (2) and  $\hat{\beta}_{1\lambda}^{LASSO} \rightarrow \hat{\beta}_1^{OLS}$ . The Lasso estimate will deviate from the OLS estimate for positive values of  $\lambda$ . Finally,  $\hat{\beta}_{1\lambda}^{LASSO}$  will be shrunk to zero as  $\lambda \rightarrow \infty$ . Therefore, for values  $\lambda \geq 0$ , the Lasso is biased towards zero if compared with OLS.

The factor  $\lambda$  is introduced for two reasons. First, the shrinkage penalty  $\sum_{s=1}^K |\beta_{s1}|$  in Lasso provides corner solutions, which implies that some coefficients are forced to be zero. Therefore, the Lasso works well for model selection when the number of candidate variables  $K$  is large. Second, for appropriate values of  $\lambda$ , the bias introduced is compensated by a reduction of variance.

In this paper, the shrinkage factor  $\lambda$  is selected with a 10-fold cross-validation algorithm,<sup>9</sup> which is a method to test the out of sample fit of the income model.<sup>10</sup> The algorithm randomly divides the first-round of data into 10 equal sized folds. By leaving one fold out (the test fold), the model is fit in the other 9 folds (the training folds). Once the income model is estimated, the withheld fold is used to predict the model. This is repeated 10 times until all folds have been left out and all observations have a predicted value. The value of  $\lambda$  is selected so that it minimizes the mean squared error (MSE) defined as  $\sum_{i=1}^{N_1} (y_{i1}^1 - \hat{y}_{i1}^1)^2 / N_1$ .

The Lasso prediction of the first-round incomes for households surveyed in the second round is

$$\hat{y}_{i1\lambda}^{2LASSO} = \hat{\beta}_{1\lambda}^{LASSO} x_{i1}^2 \quad (8)$$

Once incomes in first round are predicted for every observation in second round, we can compute the joint probability of a household  $i$  surveyed in round 2 of being poor in round 1 and escape poverty in round 2,  $\Pr(\hat{y}_{i1\lambda}^{2LASSO} < z \text{ and } y_{i2}^2 > z)$ , as well as its income change between both periods  $\Delta y_{i\lambda}^{2LASSO} = y_{i2}^2 - \hat{y}_{i1\lambda}^{2LASSO}$ .

It is important to note that this approach has several advantages with respect to previous methods. First, residuals are not used and therefore no assumption for the distribution of error terms is required. Second, and connected to the previous point, the approach described in this paper does not introduce any arbitrary underlying weight  $\gamma$  as in the non-parametric point estimate and

<sup>9</sup> This is the first stage of the cross-validation process; a second stage is explained in [section 3](#).

<sup>10</sup> Variables are standardized to have a mean of zero and standard deviation of one.

it does not require the estimation of the age-cohort correlation of residuals from cross-sections as in the parametric approach. Third, unlike the parametric approach, the method obtains household-level income changes and not just probabilities of poverty mobility.

### **3. Data, empirical approach, and a second-stage cross-validation process**

To validate the approach, this paper uses a panel subsample of the [SEDLAC](#) harmonized micro database for Peru.<sup>11</sup> The [SEDLAC](#) project consists of more than 400 household surveys in more than 25 LAC countries. This harmonization process is a joint effort of 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. The main objective of the [SEDLAC](#) dataset is to improve the access to socio-economic statistics that are comparable over time and across countries, including poverty, inequality, employment, education, social programs, among others. The harmonized panel subsample for Peru used in this paper includes households surveyed in a five-year period interval from 2007 to 2011.

Validations are done for household-level income changes as well as for poverty dynamics defined as the proportion of individuals with a harmonized per capita income lower than a US\$4 per person per day poverty line, both in 2005 purchasing power parity (PPP) per day. Income dynamics are estimated by comparing, in the second round of data, the first round predicted household per capita income obtained from applying the machine learning method described in [section 2](#) and the actual second round household per capita income. Following previous studies, the key time invariance assumption is maintained by considering only those households whose heads are between 25 and 65 years of age in all estimates so that life cycle events are avoided in general. All validations are done for two periods in time; the first period covers one year from 2010 to 2011 (4,624 households), while the second is the whole five-year period (819 households).

Following [Cruces et al. \(2015\)](#), a second stage cross-validation is considered by randomly splitting the panel dataset into two subsamples and treating each subsample as a cross-section. Therefore, the coefficients are estimated in one of these subsamples in the first round of data and applied to the second subsample in the second round. By treating each subsample of the panel

---

<sup>11</sup> See [Bourguignon \(2015\)](#) and [Gasparini, Cicowiez, and Escudero \(2013\)](#) for a description of the [SEDLAC](#) data.

dataset as a cross-section, this second stage cross-validation avoids any bias that might arise from using the panel dataset to validate the method.

This paper follows the literature to estimate income mobility by including time invariant, deterministic, and/or retrospective regressors in the underlying models. However, unlike most of the previous analysis using Synthetic panels, the harmonized data used in this paper allow to validate poverty transitions using the same underlying harmonized variables frequently used in many regional studies (e.g., [Ferreira et al. 2012](#); [Vakis et al. 2016](#)). The underlying models in this paper include a set of variables that are commonly found in surveys to ensure comparability among countries and over time. The models consider the log of per capita household income in 2005 PPP/day as the left-hand side variable and the following 39 regressors: [1] household head age, age squared, gender, and years of education; [2] regional fixed effects (Lima, Sierra Urbana, Sierra Rural, Selva Urbana, Selva Rural, Costa Urbana, and Costa Rural); and [3] the interaction between the first and the second set of covariates.<sup>12</sup>

## 4. Validation results

### *4.1 Lasso coefficients and poverty rate prediction in the first round*

The Lasso approach has at least two advantages over the OLS regression. The first advantage is related to the bias-variance trade-off; the Lasso approach shrinks the coefficients towards zero, introducing a bias that is compensated with a reduction of variance for an optimal value of  $\lambda$ . Second, since the Lasso approach produces corner solutions, it selects a subset of covariates by potentially forcing some coefficients to be zero.

The selection of the optimal shrinkage factor  $\lambda$  is shown in [Figure 1](#). The factor is chosen with a 10-fold cross-validation algorithm. The solid line and the left vertical axis show the MSE, while the dashed line and the right vertical axis present the number of non-zero coefficients. The horizontal axis in the figure presents the value of the shrinkage factor  $\lambda$ . The value of  $\lambda = 0$  corresponds to the OLS estimation, where the variance is high but the bias is zero. As  $\lambda$  increases, the variance decreases rapidly, while the bias increases at a slower pace, leading to a sharp reduction of the MSE. The lowest MSE is obtained for the values of  $\lambda$  corresponding to the dashed

---

<sup>12</sup> We do not use sampling weights in all the estimations and predictions in this paper.

vertical lines. Beyond this point, the increase in the bias more than compensates the reduction in the variance, which leads to an increase in MSE.

Figure 1 also shows that the value of non-zero coefficients drops sharply; the model corresponding to the optimal value of  $\lambda$  considers 19 non-regressors (out of 39 in total) for both the 2007-2011 and 2010-2011 periods. Figure 2 shows the variables included in models for different values of the “shrinkage” factor  $\lambda$ . Gray cells represent the variables selected. Each row in the figure represents the covariates included in the estimations, while each column represents a different value of  $\lambda$ . Four values of  $\lambda$  are considered: column [1] shows the coefficients for the OLS estimation ( $\lambda = 0$ ); column [2] presents the selected coefficients for a value of  $\lambda$  corresponding to point A in Figure 1; column [3] shows the corresponding non-zero coefficients for a value of  $\lambda$  corresponding to point B in Figure 1; and column [4] introduces the selected coefficients for the minimum MSE. The figure shows that the variables used in the 2007-2011 period are different from the ones used in the 2010-2011 one.<sup>13</sup> Mullainathan and Jann Spiess (2017) argue that changes in the parameters selected is one of the main reasons for not using the Lasso approach to learn about the underlying data-generating process.

Based on the estimated Lasso coefficients, a first step of the intra-generational mobility analysis can be done by comparing actual poverty rates in round 1 with the estimated ones that emerge when applying the machine learning approach suggested in section 2. Table 1 presents the poverty headcounts in the first round of data. The table compares the actual poverty estimates using the panel dataset and the predicted ones from the Lasso model estimated in Figure 1. All comparisons are made for both the 2007-2011 period in panel A and the 2010-2011 period in panel B. The table presents point estimates and the 95 percent confidence intervals between parenthesis. In general, the method works well; the actual point estimates are close to the predicted ones using the Lasso model. For instance, the confidence interval in the table shows that between 36 and 46 percent of people were poor in 2007, while about 41 percent of individuals were poor that year according to the Lasso approach. The method performs the least well in panel B of the table.

---

<sup>13</sup> This is also true for the different cross validation folds in each period.

#### 4.2 Joint and conditional probabilities of poverty/non-poverty transitions

The main objective of the paper is to estimate the dynamics into and out of poverty experienced by a group of individuals between two periods of time. [Table 2](#) shows the point estimates and the 95 percent confidence intervals for both the actual poverty mobility from panel data and the Lasso model approach. Comparisons are made for four joint probabilities: the probability of being poor in both rounds of data, escaping poverty, becoming poor, and remaining non-poor. Estimates are made for both the 2007-2011 period in panel A and the 2010-2011 period in panel B.

The approach performs well in general; with few exception, most of the point estimates of mobility arising from the Lasso approach fall within the 95 percent confidence interval of actual mobility from panel data. For instance, the confidence interval in the table shows that between 2 and 6 percent of people entered poverty between 2007 and 2011, while about 4 percent of individuals entered poverty according to the Lasso approach. The table also suggest that the method performs well irrespective of the length of the period.

What proportion of the initial poor escaped poverty and what proportion of the initial non-poor entered it? [Table 3](#) presents two conditional probabilities: (i) the proportion of initial poor who escaped poverty in the second round—given by  $P(y_{i2}^2 > z \mid \hat{y}_{i1\lambda}^{2LASSO} < z)$ —; and (ii) the proportion of initial non-poor who became poor between both periods—given by  $P(y_{i2}^2 < z \mid \hat{y}_{i1\lambda}^{2LASSO} > z)$ . Estimates are presented for both periods: 2007-2011 in panel A and 2010-2011 period in panel B. Results are less accurate given that both numerator and denominator in the ratios of the conditional probabilities are estimated ([Dang and Lanjouw 2013](#)).

Once again, the approach performs well; most of the point estimates from the Lasso model fall within the 95 percent confidence interval from actual panel data. For example, actual panel shows that between 4 and 11 percent of the initial non-poor fell into poverty between 2007 and 2011, while the Lasso model predicts that about 7 percent of the initial non-poor became poor between both years.

#### 4.3 Sub-group joint probabilities

How well does the approach perform in measuring poverty dynamics for subgroups of the total population? [Figures 3](#) and [4](#) validate results by estimating the joint probabilities of poverty mobility for 17 sub-groups based on the region of residence (Lima, Sierra Urbana, Sierra Rural,

Selva Urbana, Selva Rural, Costa Urbana, and Costa Rural), age (25 to 35, 36 to 45, 46 to 55, and 56 to 65 years old), gender (male or female), and education of the household head (no education, 1 to 7 years of education, 8 to 12 years of education, and more than 12 years of education). These figures compare the Lasso poverty profiles in the vertical axis with the actual panel estimates in the horizontal axis. All sub-group probabilities are based on parameters estimated for the entire population using the 10-fold cross-validation algorithm in [Figure 1](#). The approach performs well in general for estimating poverty profiles; estimates are close to the 45-degree line for almost all subgroups, regardless of the length of the period under analysis.

#### 4.4 Sub-group income growth

Another relevant question is whether this approach works well at predicting income growth— $\Delta y_{i\lambda}^{2LASSO} = y_{i2}^2 - \hat{y}_{i1\lambda}^{2LASSO}$ —for different sub-groups of the population. [Figure 5](#) validates the methodology for estimating household per capita income growth for two groups of the population defined by: (i) the dynamic poverty transitions and (ii) the quintiles of the income distribution in the second round—i.e., the non-anonymous growth incidence curves (GIC).<sup>14</sup> All estimates from the Lasso approach are compared with the actual income growth from panel data. The figure presents both the point estimate, as well as the 95 percent confidence interval.

All estimates are generally good for both sub-groups of poverty dynamics and quintiles of the income distribution. With few exceptions, Lasso estimates are close to—and fall within the 95% confidence intervals of—actual mobility for most of the cases. This is a relevant result; unlike the parametric Synthetic panel approach developed by [Dang and Lanjouw \(2013\)](#), this figure suggests that the Lasso approach performs well at predicting income growth instead of just joint probabilities of poverty transition into and out of poverty.

#### 4.5 A matching framework to improve Lasso predictions

Results in [Figure 5](#) are sufficiently encouraging to predict income growth for different sub-groups of the population between two periods of time. However, some cases can be substantially improved, especially at the two ends of the income distribution. For instance, while incomes increased for those who remained poor between 2010 and 2011, the Lasso approach predicts a

---

<sup>14</sup> As opposed to the anonymous GIC, which refer to quantile-level (or any other percentile) income growth by quantile (or any other percentile) of the income distribution ([Ravallion and Chen 2003](#)).

negative income growth for this group of individuals between the two periods—and 95% confidence intervals do not overlap.

To improve income predictions in round 1, this section introduces a variant of the initial Lasso approach in which first-round observed cross-sectional income data are matched with the first round Lasso income predictions. To do so, a random draw from round 1 of the observed empirical income distribution is assigned to each household surveyed in round 2. These values are assigned based on the position of the household in the distribution of predicted income that results from the Lasso regularization approach described in this paper. The following 4 steps describe the approach

[1] For each household in round 1, take a random draw with replacement of size  $N_2$ —which indexes the number of observations in round 2—from the empirical income distribution of actual log incomes and denote it by  $\tilde{y}_{i1}^1$ .

[2] Sort the two vectors of log incomes  $\tilde{y}_{i1}^1$  and  $\hat{y}_{i1\lambda}^{2LASSO}$  from the lowest to the highest value

$$\tilde{y}_{11}^1 \leq \tilde{y}_{21}^1 \leq \dots \leq \tilde{y}_{N_2 1}^1$$

And (9)

$$\hat{y}_{11\lambda}^{2LASSO} \leq \hat{y}_{21\lambda}^{2LASSO} \leq \dots \leq \hat{y}_{N_2 1\lambda}^{2LASSO}$$

[3] For every household in round 2, and based on the position they have in distribution of the predicted income  $\hat{y}_{i1\lambda}^{2LASSO}$ , match the two vectors of log incomes and replace the Lasso income predictions with the corresponding income from the first round.

[4] The joint probability of a household  $i$  surveyed in round 2 of being poor in round 1 and escape poverty in round 2 is given by  $\Pr(\tilde{y}_{i1}^2 < z \text{ and } y_{i2}^2 > z)$ , where  $\tilde{y}_{i1}^2$  is first round log income of household  $i$  surveyed in round 2 that results from implementing step [3]. Similarly, the change in incomes between both periods is  $\Delta\tilde{y}_i^2 = y_{i2}^2 - \tilde{y}_{i1}^2$ .

Since  $\tilde{y}_{i1}^2$  constitutes a random sample from the empirical distribution of first-round actual incomes, this matching framework is expected to outperform the Lasso predictions described in previous sections. [Table 4](#) presents all the estimates and the 95% confidence intervals based on this matching framework for the 2007-2011 and 2010-2011 periods. Panel A presents the poverty headcount in the first round of data, panel B shows the four joint probabilities of poverty mobility,

and panel C introduces the two conditional probabilities. Performance is similar to the ones observed in previous tables; most of the point estimates of mobility in [Table 4](#) fall within the 95 percent confidence interval of actual mobility from panel data.

However, results improve substantially when comparing changes in household incomes  $\Delta\tilde{y}_i^2$ . [Figure 6](#) validates this matching framework by estimating household per capita income growth for the same two groups of the population defined in [Figure 5](#). Results show a marked improvement; except for the fifth quintile, all estimates are close to and fall within the 95% confidence intervals of actual mobility.

## 5. Conclusion

This is the first paper, to the best of my knowledge, that uses a supervised machine learning approach to estimate welfare dynamics in the absence of panel datasets. It proposes to estimate parameters of a log income model in the first round of cross-sectional data using a Lasso process and use those parameters to predict incomes in the first round for all households surveyed in the second round of data. The proposed approach is validated by comparing income dynamics estimated from cross-sectional data with those derived from panel data from Peru. A validation process is implemented in two stages. In a first stage, a 10-fold cross-validation algorithm is used to evaluate the out-of-sample performance of the underlying income models in the first round of data. In a second stage, a cross-validation is implemented by randomly splitting the panel dataset into two subsamples to treat each subsample as a cross-section, which avoids any bias from using actual panel data to validate the method proposed in this paper.

A critical reason for using the approach suggested in this paper is that most of the data used to monitor poverty trends are not longitudinal in the sense that they do not follow individuals or households over time. There has been a rapid expansion in the number of household surveys in recent years, although most of these datasets are cross-sectional in nature. Panel datasets, when available, typically cover short periods of time, which poses serious concerns regarding the validity of policy recommendations that arise from their use in the analysis of long-term poverty dynamics ([Ferreira et al. 2012](#)). The proposed approach allows the analysis of poverty dynamics by describing the gross flow of household movements over time, as opposed to the net changes in poverty. This analysis helps to understand, for example, how much income mobility there has been, who has benefited from that mobility, and what have been the factors behind this mobility.

Results in this paper suggest that the method performs well in predicting the joint and conditional probabilities of entering and exiting poverty; most poverty transition estimates using cross sections fall within the 95 percent confidence intervals of mobility from panel data. The method also allows estimating household-level income growth between two periods of time in the absence of longitudinal data.

The machine learning approach introduced in this paper presents several strengths and uses less restrictive assumptions than previously developed Synthetic panel methods. As such, it serves as a promising contribution to guide future research on intra-generational income mobility. For instance, future research could expand the approach to more than two periods and/or two or more poverty lines; and consider other dependent variables (e.g., labor or health as suggested by [Dang and Lanjouw 2013](#)). Additional research could also focus on the application of this method to general situations in which two moments in time are considered, for instance, to estimate vulnerability lines based on the population at risk of falling into poverty ([Dang and Lanjouw 2016](#)).

Estimates in this paper are computed based on harmonized micro data that allow validations of poverty dynamics using the same variables frequently included in regional and global poverty analysis. The models used in the study include variables that are easy to find in all countries, which ensures the comparability of estimates between countries and over time. However, if the objective is to study income dynamics in one country—as opposed to many countries or a region as a whole—more predictive power may be achieved by including variables available in that country, but not necessarily in other countries, such as parent’s education, place of birth, etc.

This paper suggests using this machine learning approach in the absence of longitudinal data that follow individuals or households over two or more moments in time. However, the approach is not intended to be a substitute—but rather a complement—of panel data. For instance, the method can be used to combine a small panel data set with mobility estimates using this method on a larger cross-sectional data set ([Dang et al. 2014](#)) or to correct for serious non-random attrition in actual panel data sets ([Dang and Lanjouw 2013](#)).

## References

- Afzal, Marium, Jonathan Hersh, and David Newhouse. 2015. "Building a Better Model: Variable Selection to Predict Poverty in Pakistan and Sri Lanka." Mimeo, World Bank.
- Babenko, Boris, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, and Tom Swartz. 2017. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico." *Proceedings of the Neural Information Processing Systems*.
- Bourguignon, F. 2015. "Appraising income inequality databases in Latin America" *The Journal of Economic Inequality* 13 (4): 557–578.
- Castaneda Aguilar, Raul Andres; Gasparini, Leonardo Carlos; Garriga, Santiago; Lucchetti, Leonardo Ramiro; Valderrama Gonzalez, Daniel. (Forthcoming). "Measuring poverty in Latin America and the Caribbean: methodological considerations when estimating an empirical regional poverty line." *Economia Journal*. The Latin American and Caribbean Economic Association - LACEA
- CEDLAS, and World Bank. 2015. "SEDLAC: Socio-Economic Database for Latin America and the Caribbean." SEDLAC. August. <http://sedlac.econo.unlp.edu.ar/eng/>.
- Chen, Shaohua, and Martin Ravallion. 2001. "How Did the World's Poorest Fare in the 1990s?" *The Review of Income and Wealth* 47 (3): 283–300. doi:10.1111/1475-4991.00018.
- 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.
- 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.
- Dang, Hai-Anh and Peter Lanjouw. 2017. "Welfare Dynamics Measurement: Two Definitions of a Vulnerability Line and Their Empirical Applications." *The Review of Income and Wealth*. 63 (4): 633-660.

- Dang, Hai-Anh and Peter Lanjouw (forthcoming). Poverty dynamics in India between 2004-2012: Insights from longitudinal analysis using synthetic panel data. *Economic Development and Cultural Change*.
- Dang, Hai-Anh and Andrew L. Dabalen (forthcoming). Is Poverty in Africa Mostly Chronic or Transient? Evidence from Synthetic Panel Data. *Journal of Development Studies*.
- Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–64.
- Engstrom, Ryan; Hersh, Jonathan Samuel; Newhouse, David Locke. 2017. Poverty from space: using high-resolution satellite imagery for estimating economic well-being (English). Policy Research working paper; no. WPS 8284. Washington, D.C.: World Bank Group.
- 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.
- Ferreira, Francisco H. G., Shaohua Chen, Andrew L. Dabalen, Yuri M. Dikhanov, Nada Hamadeh, Dean Mitchell Jolliffe, Ambar Narayan, et al. 2016. "A Global Count of the Extreme Poor in 2012: Data Issues, Methodology and Initial Results." *The Journal of Economic Inequality* 14 (2): 141–72.
- 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.
- Jolliffe, Dean, and Espen Beer Prydz. 2016. "Estimating International Poverty Lines from Comparable National Thresholds." *The Journal of Economic Inequality* 14 (2): 185–98.
- Lucchetti, Leonardo. 2017. "Who Escaped Poverty and Who Was Left Behind? A Non-Parametric Approach to Explore Welfare Dynamics Using Cross-Sections." World Bank Policy Research Working Paper No. 8220.
- McBride, Linden; Nichols, Austin. 2016. "Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning." *World Bank Economic Review*, lhw056.
- Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives*, 31 (2): 87-106.
- Ravallion, Martin, Shaohua Chen, and Prem Sangraula. 2009. "Dollar a Day Revisited." *The World Bank Economic Review*, June, lhp007.
- Ravallion, M., and S. Chen. 2003. "Measuring Propoor Growth." *Economics Letters* 78 (1): 93–99.

Ravallion, Martin, Gaurav Datt, and Dominique van de Walle. 1991. "Quantifying Absolute Poverty in the Developing World." *Review of Income and Wealth* 37 (4): 345–61.

Serajuddin, Umar; Uematsu, Hiroki; Wieser, Christina; Yoshida, Nobuo; Dabalen, Andrew L. 2015. "Data deprivation: another deprivation to end." Policy Research working paper; no. WPS 7252. Washington, D.C.: World Bank Group.

Tibshirani, R. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

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 headcounts in the first round using 2011 observations

| Status in Round 1         | Actual         | LASSO          |
|---------------------------|----------------|----------------|
|                           | [1]            | [2]            |
| <i>Panel A: Peru 2007</i> |                |                |
| Poverty Rate              | 42<br>(36, 46) | 41<br>(35, 45) |
| <i>Panel B: Peru 2010</i> |                |                |
| Poverty Rate              | 30<br>(27, 31) | 26<br>(23, 27) |
| Obs. Panel A              | 409            | 409            |
| Obs. Panel B              | 2,312          | 2,312          |

Data source: [SEDLAC](#) data (CEDLAS and the World Bank). 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 estimates. Column [2] shows Machine Learning estimates. Poor are those individuals with a per capita income lower than \$4 a day. Poverty lines and incomes are expressed in 2005 \$PPP/day. 95% confidence intervals between parenthesis. All results are unweighted.

Table 2. Transition matrices – actual panel data and Lasso estimates using repeated cross sections and the 2011 observations  
*Unconditional probability*

| Status in t=1,2                | Actual         | LASSO          |
|--------------------------------|----------------|----------------|
|                                | [1]            | [2]            |
| <i>Panel A: Peru 2007-2011</i> |                |                |
| Poor, Poor                     | 23<br>(18, 27) | 23<br>(19, 27) |
| Poor, Non-poor                 | 19<br>(14, 22) | 17<br>(13, 20) |
| Non-poor, Poor                 | 5<br>(2, 6)    | 4<br>(2, 6)    |
| Non-poor, Non-poor             | 54<br>(48, 58) | 55<br>(50, 59) |
| <i>Panel B: Peru 2010-2011</i> |                |                |
| Poor, Poor                     | 20<br>(18, 21) | 17<br>(15, 18) |
| Poor, Non-poor                 | 10<br>(8, 10)  | 9<br>(7, 10)   |
| Non-poor, Poor                 | 8<br>(6, 9)    | 11<br>(10, 12) |
| Non-poor, Non-poor             | 62<br>(60, 64) | 63<br>(61, 65) |
| Observations panel A           | 409            | 409            |
| Observations panel B           | 2,312          | 2,312          |

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. Results in column [1] show actual panel mobility. Column [2] shows Machine Learning estimates. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. 95% confidence intervals between parenthesis. All results are unweighted.

Table 3. Transition matrices – actual panel data and Lasso estimates using repeated cross sections and the 2011 observations  
*Conditional probability*

| Conditional Mobility  | Actual<br>[1]  | LASSO<br>[2]   |
|---|----------------|----------------|
| <i>Panel A: Peru 2007 - 2011</i>                              |                |                |
| Proportion of poor in 2007 who moved out of poverty in 2011   | 45<br>(37, 52) | 42<br>(34, 49) |
| Proportion of non-poor in 2007 who moved into poverty in 2011 | 8<br>(4, 11)   | 7<br>(4, 10)   |
| <i>Panel B: Peru 2010 - 2011</i>                              |                |                |
| Proportion of poor in 2010 who moved out of poverty in 2011   | 33<br>(29, 36) | 35<br>(30, 38) |
| Proportion of non-poor in 2010 who moved into poverty in 2011 | 11<br>(9, 12)  | 15<br>(13, 16) |
| Observations Panel A  | 409            | 409            |
| Observations Panel B  | 2,312          | 2,312          |

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. Column [1] shows actual panel mobility. Column [2] shows Machine Learning estimates. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. 95% confidence intervals between parenthesis. All results are unweighted.

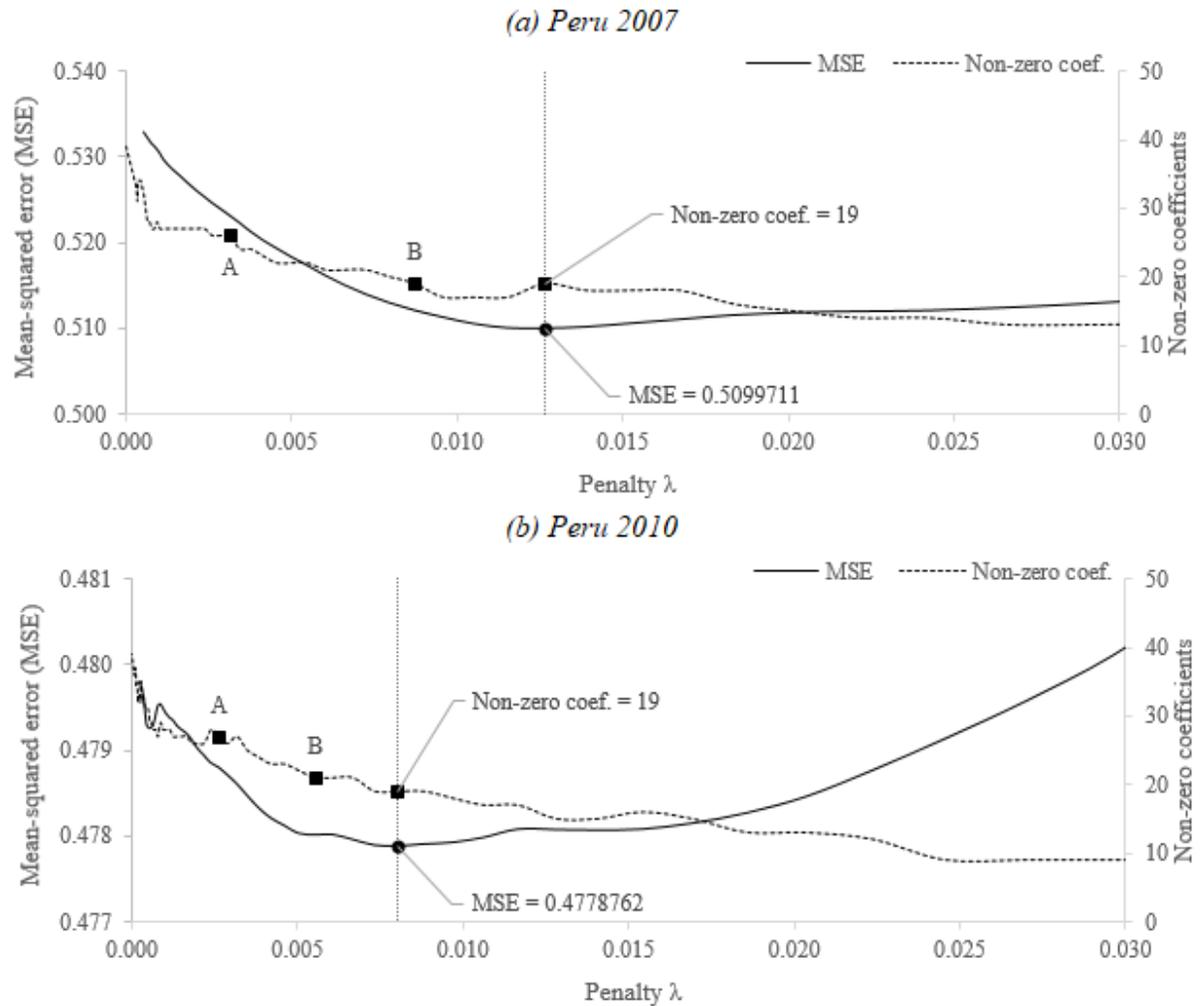
Table 4: Simulated poverty in the first round and transition matrices using 2011 observations

*Matching first round cross-sectional data and LASSO predictions*

| Poverty level and transition   | Panel 2007-2011<br>[1] | Panel 2010-2011<br>[2] |
|--|------------------------|------------------------|
| <i>Panel A: Poverty in first round</i>                               |                        |                        |
| Poverty Rate   | 46<br>(41, 50)         | 30<br>(27, 31)         |
| <i>Panel B: Unconditional probabilities</i>                          |                        |                        |
| Poor, Poor   | 24<br>(20, 28)         | 19<br>(17, 20)         |
| Poor, Non-poor   | 22<br>(17, 25)         | 11<br>(9, 12)          |
| Non-poor, Poor   | 3<br>(1, 5)            | 9<br>(8, 10)           |
| Non-poor, Non-poor   | 51<br>(45, 55)         | 61<br>(58, 62)         |
| <i>Panel C: Conditional probabilities</i>                            |                        |                        |
| Proportion of poor in first round who moved out of poverty in 2011   | 47<br>(39, 53)         | 37<br>(33, 40)         |
| Proportion of non-poor in first round who moved into poverty in 2011 | 6<br>(3, 9)            | 13<br>(11, 15)         |
| Observations   | 409                    | 2,312                  |

Data source: SEDLAC data (CEDLAS and the World Bank). Note: The table presents results that arise from matching first round cross-sectional incomes with LASSO predictions. Results in column [1] shows matching predictions using 2007-2011 data, while column [2] shows matching predictions using 2010-2011 data. Panel A presents poverty in the first-round data. Panel B shows unconditional probabilities of poverty transition. Panel C presents conditional probabilities of poverty transitions. Poor are those individuals with a per capita income lower than \$4 a day. Poverty lines and incomes are expressed in 2005 \$PPP/day.

Figure 1. Out-of-sample cross-validation profile for the Lasso regression model



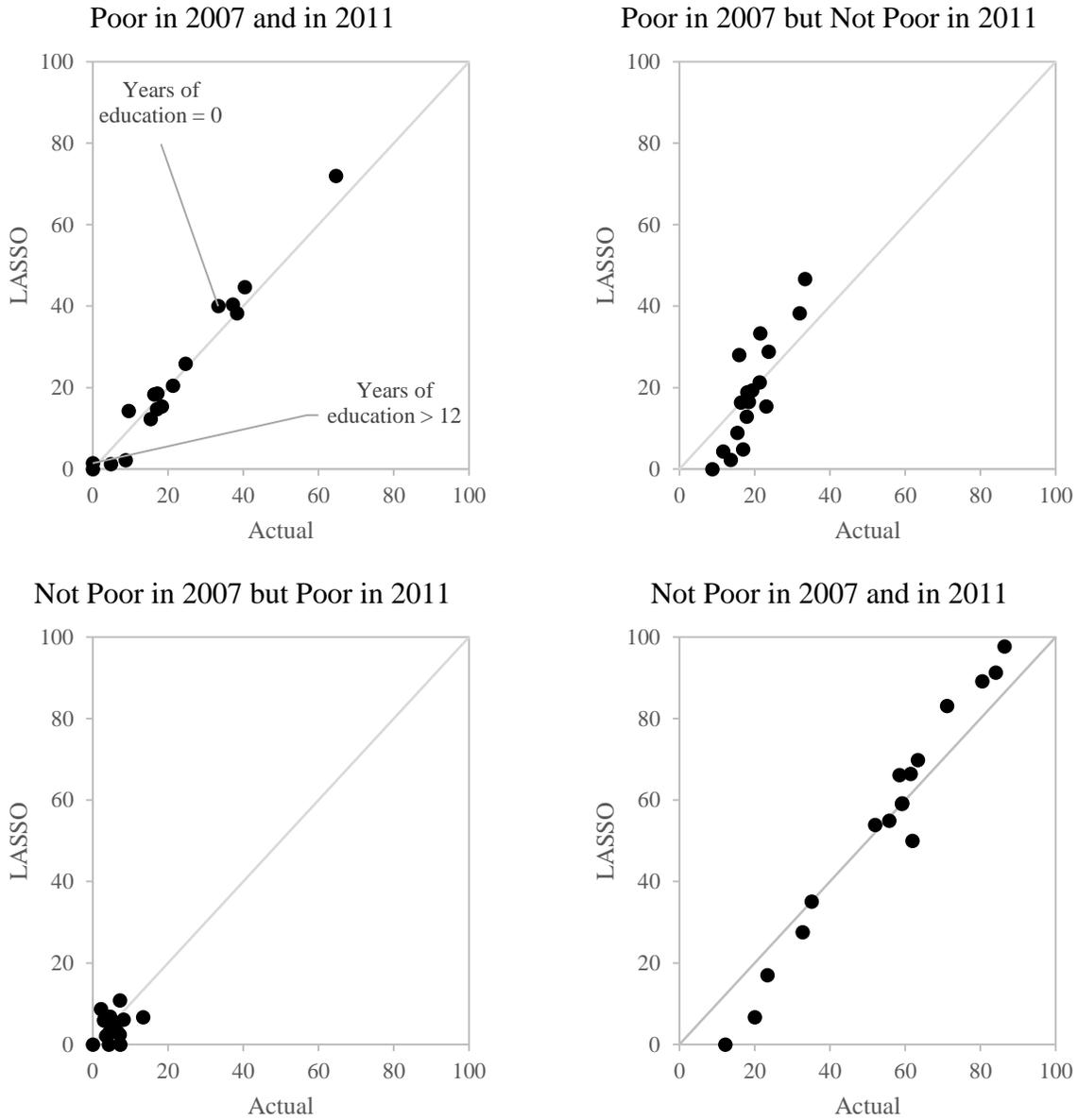
Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Incomes are expressed in 2005 \$PPP/day. All results are unweighted.

Figure 2. Non-zero Lasso coefficients for different values of  $\lambda$  using repeated cross sections and the 2011 observations

| Coefficients      |              | 2007-2011 |             |             |             | 2010-2011 |             |             |             |
|-------------------|--------------|-----------|-------------|-------------|-------------|-----------|-------------|-------------|-------------|
|                   |              | OLS [1]   | Point A [2] | Point B [3] | Min MSE [4] | OLS [1]   | Point A [2] | Point B [3] | Min MSE [4] |
| Levels            | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |
| Region Fixed Eff. | U. Costa     |           |             |             |             |           |             |             |             |
|                   | U. Sierra    |           |             |             |             |           |             |             |             |
|                   | U. Selva     |           |             |             |             |           |             |             |             |
|                   | R. Costa     |           |             |             |             |           |             |             |             |
|                   | R. Sierra    |           |             |             |             |           |             |             |             |
|                   | R. Selva     |           |             |             |             |           |             |             |             |
| Interactions      | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |
|                   | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |
|                   | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |
|                   | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |
|                   | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |
|                   | Male         |           |             |             |             |           |             |             |             |
|                   | Age          |           |             |             |             |           |             |             |             |
|                   | Age Sq.      |           |             |             |             |           |             |             |             |
|                   | Years of ed. |           |             |             |             |           |             |             |             |

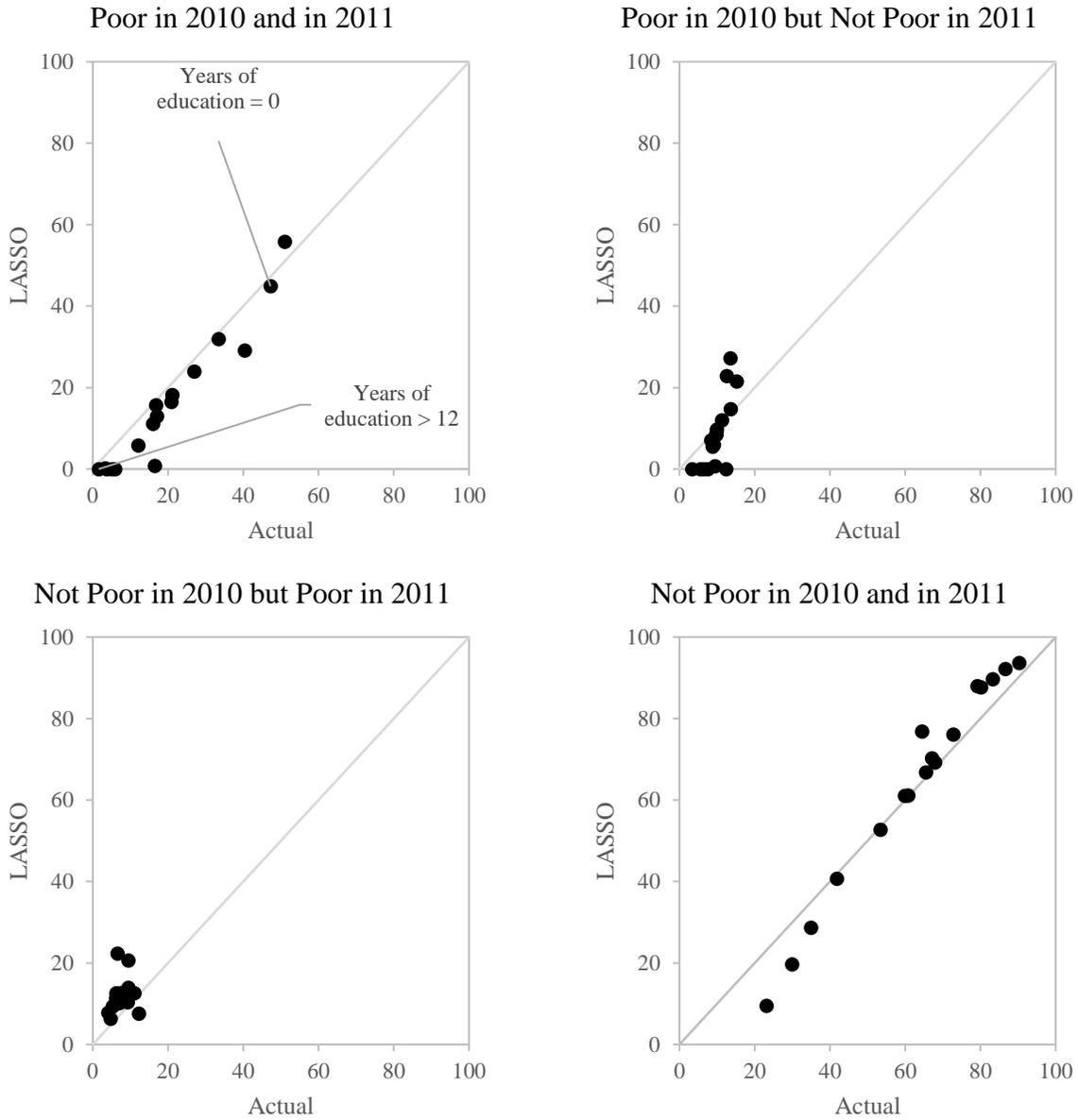
Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Each column represents a different value of  $\lambda$ . Results in column [1] shows the coefficients for  $\lambda = 0$ ; column [2] shows results for  $\lambda$  corresponding to point A in [Figure 1](#); column [3] presents results for  $\lambda$  corresponding to point B in [Figure 1](#); and column [4] presents results for  $\lambda$  corresponding to the minimum MSE. Gray cells represent the variables selected. Each row in the figure represents the covariates included. Incomes are expressed in 2005 \$PPP/day. All results are unweighted.

Figure 3. 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. 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. All results are unweighted.

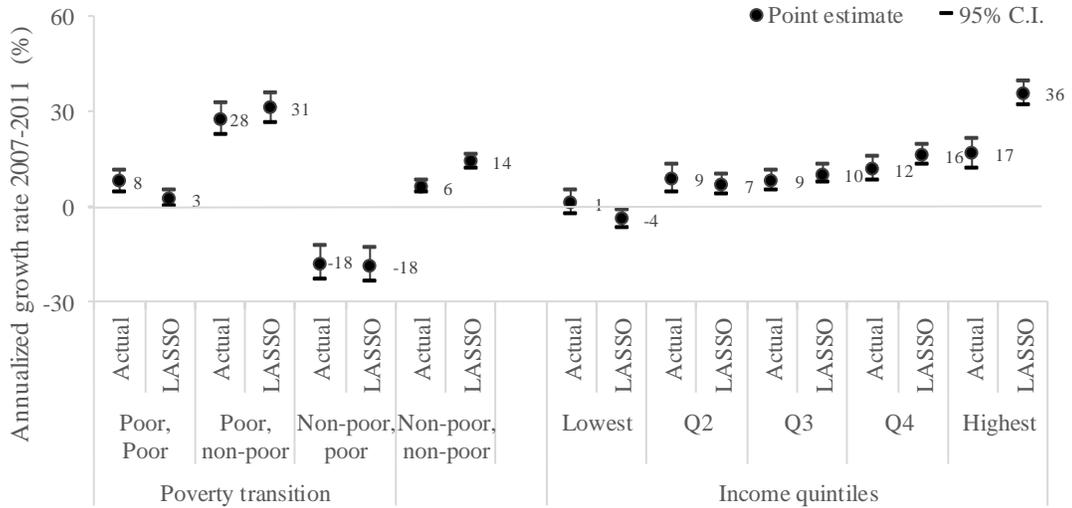
Figure 4. Poverty dynamics by subgroups of the population  
Peru 2010 and 2011



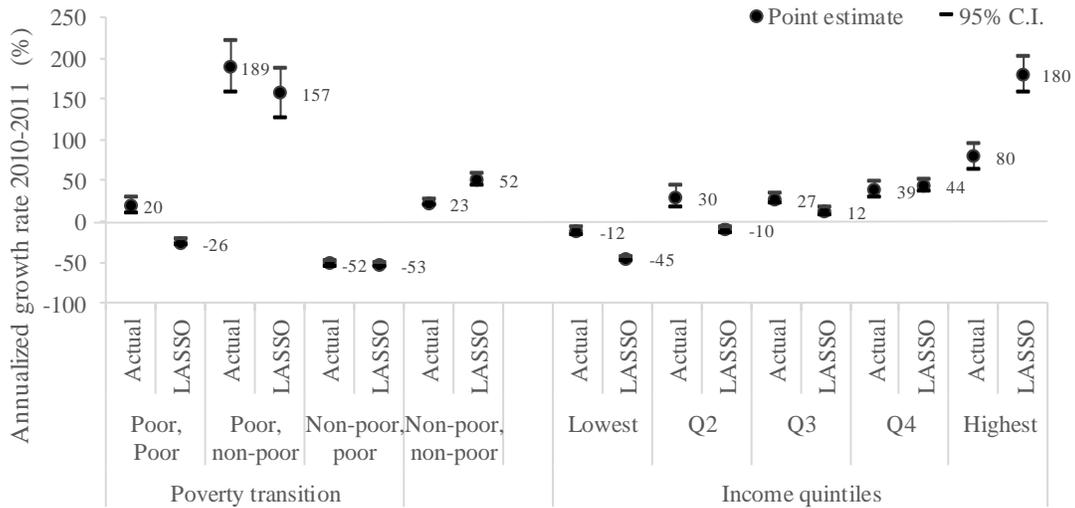
Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. The 45-degree line shows actual panel mobility. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. All results are unweighted.

Figure 5. Household-level income change by groups of mobility transition and quintiles of the income distribution

(a) Peru 2007 and 2011

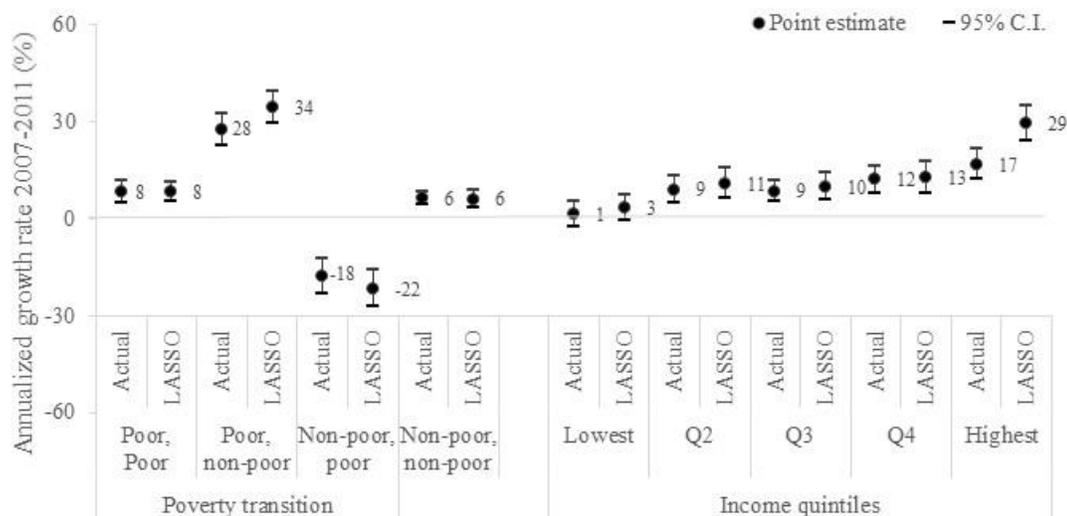


(b) Peru 2010 and 2011

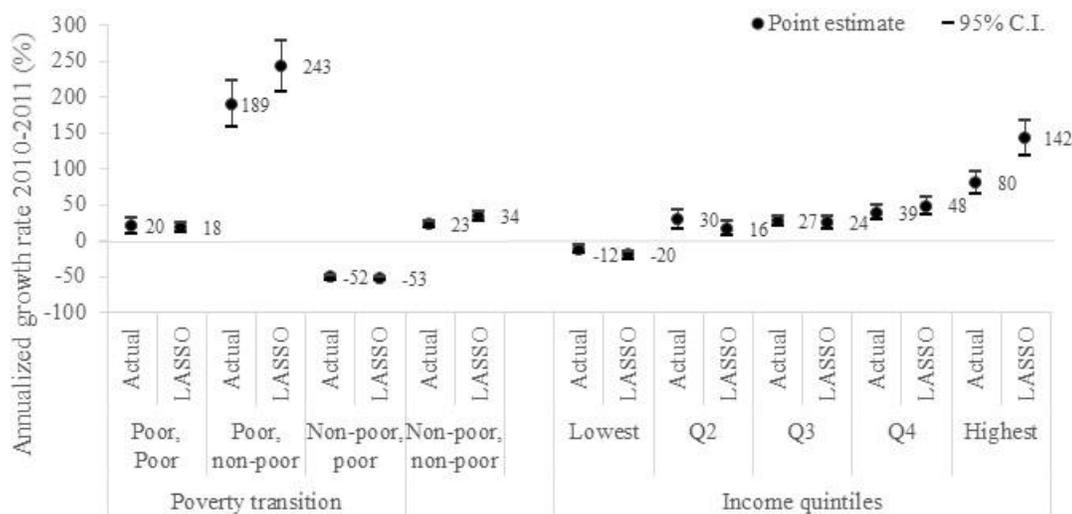


Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. All results are unweighted.

Figure 6. Household-level income change by groups of mobility transition and quintiles of the income distribution - Matching first round cross-sectional data and LASSO predictions  
 (a) Peru 2007 and 2011



(b) Peru 2010 and 2011



Data source: [SEDLAC](#) data (CEDLAS and the World Bank). Note: The figure presents results that arise from randomly drawing actual income from round 1 and allocating that income to each household surveyed in round 2 according to their position in the distribution of predicted income that results from the Lasso approach described in this paper (presented as “LASSO” in the figure), as well as estimates using “actual” data. Results are constrained to the panel sample of households whose heads are between 25 and 65 years old. Poor are those individuals with a per capita income lower than \$4. Poverty lines and incomes are expressed in 2005 \$PPP/day. All results are unweighted.



## Poverty & Equity Global Practice Working Papers (Since July 2014)

The Poverty & Equity Global Practice 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.

This series is co-published with the World Bank Policy Research Working Papers (DECOS). It is part of a larger effort by the World Bank to provide open access to its research and contribute to development policy discussions around the world.

For the latest paper, visit our GP's intranet at <http://POVERTY>.

- 1 **Estimating poverty in the absence of consumption data: the case of Liberia**  
*Dabalen, A. L., Graham, E., Himelein, K., Mungai, R.,* September 2014
- 2 **Female labor participation in the Arab world: some evidence from panel data in Morocco**  
*Barry, A. G., Guennouni, J., Verme, P.,* September 2014
- 3 **Should income inequality be reduced and who should benefit? redistributive preferences in Europe and Central Asia**  
*Cojocaru, A., Diagne, M. F.,* November 2014
- 4 **Rent imputation for welfare measurement: a review of methodologies and empirical findings**  
*Balcazar Salazar, C. F., Ceriani, L., Olivieri, S., Ranzani, M.,* November 2014
- 5 **Can agricultural households farm their way out of poverty?**  
*Oseni, G., McGee, K., Dabalen, A.,* November 2014
- 6 **Durable goods and poverty measurement**  
*Amendola, N., Vecchi, G.,* November 2014
- 7 **Inequality stagnation in Latin America in the aftermath of the global financial crisis**  
*Cord, L., Barriga Cabanillas, O., Lucchetti, L., Rodriguez-Castelan, C., Sousa, L. D., Valderrama, D.,* December 2014
- 8 **Born with a silver spoon: inequality in educational achievement across the world**  
*Balcazar Salazar, C. F., Narayan, A., Tiwari, S.,* January 2015

- 9 **Long-run effects of democracy on income inequality: evidence from repeated cross-sections**  
*Balcazar Salazar, C. F.*, January 2015
- 10 **Living on the edge: vulnerability to poverty and public transfers in Mexico**  
*Ortiz-Juarez, E., Rodriguez-Castelan, C., De La Fuente, A.*, January 2015
- 11 **Moldova: a story of upward economic mobility**  
*Davalos, M. E., Meyer, M.*, January 2015
- 12 **Broken gears: the value added of higher education on teachers' academic achievement**  
*Balcazar Salazar, C. F., Nopo, H.*, January 2015
- 13 **Can we measure resilience? a proposed method and evidence from countries in the Sahel**  
*Alfani, F., Dabalen, A. L., Fisker, P., Molini, V.*, January 2015
- 14 **Vulnerability to malnutrition in the West African Sahel**  
*Alfani, F., Dabalen, A. L., Fisker, P., Molini, V.*, January 2015
- 15 **Economic mobility in Europe and Central Asia: exploring patterns and uncovering puzzles**  
*Cancho, C., Davalos, M. E., Demarchi, G., Meyer, M., Sanchez Paramo, C.*, January 2015
- 16 **Managing risk with insurance and savings: experimental evidence for male and female farm managers in the Sahel**  
*Delavallade, C., Dizon, F., Hill, R., Petraud, J. P., et al.*, January 2015
- 17 **Gone with the storm: rainfall shocks and household well-being in Guatemala**  
*Baez, J. E., Lucchetti, L., Genoni, M. E., Salazar, M.*, January 2015
- 18 **Handling the weather: insurance, savings, and credit in West Africa**  
*De Nicola, F.*, February 2015
- 19 **The distributional impact of fiscal policy in South Africa**  
*Inchauste Comboni, M. G., Lustig, N., Maboshe, M., Purfield, C., Woolard, I.*, March 2015
- 20 **Interviewer effects in subjective survey questions: evidence from Timor-Leste**  
*Himelein, K.*, March 2015
- 21 **No condition is permanent: middle class in Nigeria in the last decade**  
*Corral Rodas, P. A., Molini, V., Oseni, G. O.*, March 2015
- 22 **An evaluation of the 2014 subsidy reforms in Morocco and a simulation of further reforms**  
*Verme, P., El Massnaoui, K.*, March 2015

- 23 **The quest for subsidy reforms in Libya**  
*Araar, A., Choueiri, N., Verme, P.,* March 2015
- 24 **The (non-) effect of violence on education: evidence from the "war on drugs" in Mexico**  
*Márquez-Padilla, F., Pérez-Arce, F., Rodríguez Castelan, C.,* April 2015
- 25 **"Missing girls" in the south Caucasus countries: trends, possible causes, and policy options**  
*Das Gupta, M.,* April 2015
- 26 **Measuring inequality from top to bottom**  
*Diaz Bazan, T. V.,* April 2015
- 27 **Are we confusing poverty with preferences?**  
*Van Den Boom, B., Halsema, A., Molini, V.,* April 2015
- 28 **Socioeconomic impact of the crisis in north Mali on displaced people** (Available in French)  
*Etang Ndip, A., Hoogeveen, J. G., Lendorfer, J.,* June 2015
- 29 **Data deprivation: another deprivation to end**  
*Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., Dabalen, A.,* April 2015
- 30 **The local socioeconomic effects of gold mining: evidence from Ghana**  
*Chuhan-Pole, P., Dabalen, A., Kotsadam, A., Sanoh, A., Tolonen, A.K.,* April 2015
- 31 **Inequality of outcomes and inequality of opportunity in Tanzania**  
*Belghith, N. B. H., Zeufack, A. G.,* May 2015
- 32 **How unfair is the inequality of wage earnings in Russia? estimates from panel data**  
*Tiwari, S., Lara Ibarra, G., Narayan, A.,* June 2015
- 33 **Fertility transition in Turkey—who is most at risk of deciding against child arrival?**  
*Greulich, A., Dasre, A., Inan, C.,* June 2015
- 34 **The socioeconomic impacts of energy reform in Tunisia: a simulation approach**  
*Cuesta Leiva, J. A., El Lahga, A., Lara Ibarra, G.,* June 2015
- 35 **Energy subsidies reform in Jordan: welfare implications of different scenarios**  
*Atamanov, A., Jellema, J. R., Serajuddin, U.,* June 2015
- 36 **How costly are labor gender gaps? estimates for the Balkans and Turkey**  
*Cuberes, D., Teignier, M.,* June 2015
- 37 **Subjective well-being across the lifespan in Europe and Central Asia**  
*Bauer, J. M., Munoz Boudet, A. M., Levin, V., Nie, P., Sousa-Poza, A.,* July 2015

- 38 **Lower bounds on inequality of opportunity and measurement error**  
*Balcazar Salazar, C. F.*, July 2015
- 39 **A decade of declining earnings inequality in the Russian Federation**  
*Posadas, J., Calvo, P. A., Lopez-Calva, L.-F.*, August 2015
- 40 **Gender gap in pay in the Russian Federation: twenty years later, still a concern**  
*Atencio, A., Posadas, J.*, August 2015
- 41 **Job opportunities along the rural-urban gradation and female labor force participation in India**  
*Chatterjee, U., Rama, M. G., Murgai, R.*, September 2015
- 42 **Multidimensional poverty in Ethiopia: changes in overlapping deprivations**  
*Yigezu, B., Ambel, A. A., Mehta, P. A.*, September 2015
- 43 **Are public libraries improving quality of education? when the provision of public goods is not enough**  
*Rodriguez Lesmes, P. A., Valderrama Gonzalez, D., Trujillo, J. D.*, September 2015
- 44 **Understanding poverty reduction in Sri Lanka: evidence from 2002 to 2012/13**  
*Inchauste Comboni, M. G., Ceriani, L., Olivieri, S. D.*, October 2015
- 45 **A global count of the extreme poor in 2012: data issues, methodology and initial results**  
*Ferreira, F.H.G., Chen, S., Dabalén, A. L., Dikhanov, Y. M., Hamadeh, N., Jolliffe, D. M., Narayan, A., Prydz, E. B., Revenga, A. L., Sangraula, P., Serajuddin, U., Yoshida, N.*, October 2015
- 46 **Exploring the sources of downward bias in measuring inequality of opportunity**  
*Lara Ibarra, G., Martínez Cruz, A. L.*, October 2015
- 47 **Women's police stations and domestic violence: evidence from Brazil**  
*Perova, E., Reynolds, S.*, November 2015
- 48 **From demographic dividend to demographic burden? regional trends of population aging in Russia**  
*Matytsin, M., Moorty, L. M., Richter, K.*, November 2015
- 49 **Hub-periphery development pattern and inclusive growth: case study of Guangdong province**  
*Luo, X., Zhu, N.*, December 2015
- 50 **Unpacking the MPI: a decomposition approach of changes in multidimensional poverty headcounts**  
*Rodriguez Castelan, C., Trujillo, J. D., Pérez Pérez, J. E., Valderrama, D.*, December 2015
- 51 **The poverty effects of market concentration**  
*Rodriguez Castelan, C.*, December 2015
- 52 **Can a small social pension promote labor force participation? evidence from the Colombia Mayor program**  
*Pfütze, T., Rodriguez Castelan, C.*, December 2015

- 53 **Why so gloomy? perceptions of economic mobility in Europe and Central Asia**  
*Davalos, M. E., Cancho, C. A., Sanchez, C.,* December 2015
- 54 **Tenure security premium in informal housing markets: a spatial hedonic analysis**  
*Nakamura, S.,* December 2015
- 55 **Earnings premiums and penalties for self-employment and informal employees around the world**  
*Newhouse, D. L., Mossaad, N., Gindling, T. H.,* January 2016
- 56 **How equitable is access to finance in turkey? evidence from the latest global FINDEX**  
*Yang, J., Azevedo, J. P. W. D., Inan, O. K.,* January 2016
- 57 **What are the impacts of Syrian refugees on host community welfare in Turkey? a subnational poverty analysis**  
*Yang, J., Azevedo, J. P. W. D., Inan, O. K.,* January 2016
- 58 **Declining wages for college-educated workers in Mexico: are younger or older cohorts hurt the most?**  
*Lustig, N., Campos-Vazquez, R. M., Lopez-Calva, L.-F.,* January 2016
- 59 **Sifting through the Data: labor markets in Haiti through a turbulent decade (2001-2012)**  
*Rodella, A.-S., Scot, T.,* February 2016
- 60 **Drought and retribution: evidence from a large-scale rainfall-indexed insurance program in Mexico**  
*Fuchs Tarlovsky, Alan., Wolff, H.,* February 2016
- 61 **Prices and welfare**  
*Verme, P., Araar, A.,* February 2016
- 62 **Losing the gains of the past: the welfare and distributional impacts of the twin crises in Iraq 2014**  
*Olivieri, S. D., Krishnan, N.,* February 2016
- 63 **Growth, urbanization, and poverty reduction in India**  
*Ravallion, M., Murgai, R., Datt, G.,* February 2016
- 64 **Why did poverty decline in India? a nonparametric decomposition exercise**  
*Murgai, R., Balcazar Salazar, C. F., Narayan, A., Desai, S.,* March 2016
- 65 **Robustness of shared prosperity estimates: how different methodological choices matter**  
*Uematsu, H., Atamanov, A., Dewina, R., Nguyen, M. C., Azevedo, J. P. W. D., Wieser, C., Yoshida, N.,* March 2016
- 66 **Is random forest a superior methodology for predicting poverty? an empirical assessment**  
*Stender, N., Pave Sohnesen, T.,* March 2016
- 67 **When do gender wage differences emerge? a study of Azerbaijan's labor market**  
*Tiongson, E. H. R., Pastore, F., Sattar, S.,* March 2016

- 68 **Second-stage sampling for conflict areas: methods and implications**  
*Eckman, S., Murray, S., Himelein, K., Bauer, J., March 2016*
- 69 **Measuring poverty in Latin America and the Caribbean: methodological considerations when estimating an empirical regional poverty line**  
*Gasparini, L. C., April 2016*
- 70 **Looking back on two decades of poverty and well-being in India**  
*Murgai, R., Narayan, A., April 2016*
- 71 **Is living in African cities expensive?**  
*Yamanaka, M., Dikhanov, Y. M., Rissanen, M. O., Harati, R., Nakamura, S., Lall, S. V., Hamadeh, N., Vigil Oliver, W., April 2016*
- 72 **Ageing and family solidarity in Europe: patterns and driving factors of intergenerational support**  
*Albertini, M., Sinha, N., May 2016*
- 73 **Crime and persistent punishment: a long-run perspective on the links between violence and chronic poverty in Mexico**  
*Rodriguez Castelan, C., Martinez-Cruz, A. L., Lucchetti, L. R., Valderrama Gonzalez, D., Castaneda Aguilar, R. A., Garriga, S., June 2016*
- 74 **Should I stay or should I go? internal migration and household welfare in Ghana**  
*Molini, V., Pavelesku, D., Ranzani, M., July 2016*
- 75 **Subsidy reforms in the Middle East and North Africa Region: a review**  
*Verme, P., July 2016*
- 76 **A comparative analysis of subsidy reforms in the Middle East and North Africa Region**  
*Verme, P., Araar, A., July 2016*
- 77 **All that glitters is not gold: polarization amid poverty reduction in Ghana**  
*Clementi, F., Molini, V., Schettino, F., July 2016*
- 78 **Vulnerability to Poverty in rural Malawi**  
*Mccarthy, N., Brubaker, J., De La Fuente, A., July 2016*
- 79 **The distributional impact of taxes and transfers in Poland**  
*Goraus Tanska, K. M., Inchauste Comboni, M. G., August 2016*
- 80 **Estimating poverty rates in target populations: an assessment of the simple poverty scorecard and alternative approaches**  
*Vinha, K., Rebolledo Dellepiane, M. A., Skoufias, E., Diamond, A., Gill, M., Xu, Y., August 2016*

- 81 **Synergies in child nutrition: interactions of food security, health and environment, and child care**  
*Skoufias, E.*, August 2016
- 82 **Understanding the dynamics of labor income inequality in Latin America**  
*Rodriguez Castelan, C., Lustig, N., Valderrama, D., Lopez-Calva, L.-F.*, August 2016
- 83 **Mobility and pathways to the middle class in Nepal**  
*Tiwari, S., Balcazar Salazar, C. F., Shidiq, A. R.*, September 2016
- 84 **Constructing robust poverty trends in the Islamic Republic of Iran: 2008-14**  
*Salehi Isfahani, D., Atamanov, A., Mostafavi, M.-H., Vishwanath, T.*, September 2016
- 85 **Who are the poor in the developing world?**  
*Newhouse, D. L., Uematsu, H., Doan, D. T. T., Nguyen, M. C., Azevedo, J. P. W. D., Castaneda Aguilar, R. A.*, October 2016
- 86 **New estimates of extreme poverty for children**  
*Newhouse, D. L., Suarez Becerra, P., Evans, M. C.*, October 2016
- 87 **Shedding light: understanding energy efficiency and electricity reliability**  
*Carranza, E., Meeks, R.*, November 2016
- 88 **Heterogeneous returns to income diversification: evidence from Nigeria**  
*Siwatu, G. O., Corral Rodas, P. A., Bertoni, E., Molini, V.*, November 2016
- 89 **How liberal is Nepal's liberal grade promotion policy?**  
*Sharma, D.*, November 2016
- 90 **Pro-growth equity: a policy framework for the twin goals**  
*Lopez-Calva, L. F., Rodriguez Castelan, C.*, November 2016
- 91 **CPI bias and its implications for poverty reduction in Africa**  
*Dabalen, A. L., Gaddis, I., Nguyen, N. T. V.*, December 2016
- 92 **Building an ex ante simulation model for estimating the capacity impact, benefit incidence, and cost effectiveness of child care subsidies: an application using provider-level data from Turkey**  
*Aran, M. A., Munoz Boudet, A., Aktakke, N.*, December 2016
- 93 **Vulnerability to drought and food price shocks: evidence from Ethiopia**  
*Porter, C., Hill, R.*, December 2016
- 94 **Job quality and poverty in Latin America**  
*Rodriguez Castelan, C., Mann, C. R., Brummund, P.*, December 2016
- 95 **With a little help: shocks, agricultural income, and welfare in Uganda**  
*Mejia-Mantilla, C., Hill, R.*, January 2017

- 96 **The impact of fiscal policy on inequality and poverty in Chile**  
*Martinez Aguilar, S. N., Fuchs Tarlovsky, A., Ortiz-Juarez, E., Del Carmen Hasbun, G. E., January 2017*
- 97 **Conditionality as targeting? participation and distributional effects of conditional cash transfers**  
*Rodriguez Castelan, C., January 2017*
- 98 **How is the slowdown affecting households in Latin America and the Caribbean?**  
*Reyes, G. J., Calvo-Gonzalez, O., Sousa, L. D. C., Castaneda Aguilar, R. A., Farfan Bertran, M. G., January 2017*
- 99 **Are tobacco taxes really regressive? evidence from Chile**  
*Fuchs Tarlovsky, A., Meneses, F. J., March 2017*
- 100 **Design of a multi-stage stratified sample for poverty and welfare monitoring with multiple objectives: a Bangladesh case study**  
*Yanez Pagans, M., Roy, D., Yoshida, N., Ahmed, F., March 2017*
- 101 **For India's rural poor, growing towns matter more than growing cities**  
*Murgai, R., Ravallion, M., Datt, G., Gibson, J., March 2017*
- 102 **Leaving, staying, or coming back? migration decisions during the northern Mali conflict**  
*Hoogeveen, J. G., Sansone, D., Rossi, M., March 2017*
- 103 **Arithmetics and Politics of Domestic Resource Mobilization**  
*Bolch, K. B., Ceriani, L., Lopez-Calva, L.-F., April 2017*
- 104 **Can Public Works Programs Reduce Youth Crime? Evidence from Papua New Guinea's Urban Youth Employment Project**  
*Oleksiy I., Darian N., David N., Sonya S., April 2017*
- 105 **Is Poverty in Africa Mostly Chronic or Transient? Evidence from Synthetic Panel Data**  
*Dang, H.-A. H., Dabalén, A. L., April 2017*
- 106 **To Sew or Not to Sew? Assessing the Welfare Effects of the Garment Industry in Cambodia**  
*Mejía-Mantilla, C., Woldemichael, M. T., May 2017*
- 107 **Perceptions of distributive justice in Latin America during a period of falling inequality**  
*Reyes, G. J., Gasparini, L. C., May 2017*
- 108 **How do women fare in rural non-farm economy?**  
*Fuje, H. N., May 2017*
- 109 **Rural Non-Farm Employment and Household Welfare: Evidence from Malawi**  
*Adjognon, G. S., Liverpool-Tasie, S. L., De La Fuente, A., Benfica, R. M., May 2017*

- 110 **Multidimensional Poverty in the Philippines, 2004-13: Do Choices for Weighting, Identification and Aggregation Matter?**  
*Datt, G.*, June 2017
- 111 **But ... what is the poverty rate today? testing poverty nowcasting methods in Latin America and the Caribbean**  
*Caruso, G. D., Lucchetti, L. R., Malasquez, E., Scot, T., Castaneda, R. A.*, June 2017
- 112 **Estimating the Welfare Costs of Reforming the Iraq Public Distribution System: A Mixed Demand Approach**  
*Krishnan, N., Olivieri, S., Ramadan, R.*, June 2017
- 113 **Beyond Income Poverty: Nonmonetary Dimensions of Poverty in Uganda**  
*Etang Ndip, A., Tsimpo, C.*, June 2017
- 114 **Education and Health Services in Uganda: Quality of Inputs, User Satisfaction, and Community Welfare Levels**  
*Tsimpo Nkengne, C., Etang Ndip, A., Wodon, Q. T.*, June 2017
- 115 **Rental Regulation and Its Consequences on Measures of Well-Being in the Arab Republic of Egypt**  
*Lara Ibarra, G., Mendiratta, V., Vishwanath, T.*, July 2017
- 116 **The Poverty Implications of Alternative Tax Reforms: Results from a Numerical Application to Pakistan**  
*Feltenstein, A., Mejia-Mantilla, C., Newhouse, D. L., Sedrakyan, G.*, August 2017
- 117 **Tracing Back the Weather Origins of Human Welfare: Evidence from Mozambique?**  
*Baez Ramirez, J. E., Caruso, G. D., Niu, C.*, August 2017
- 118 **Many Faces of Deprivation: A multidimensional approach to poverty in Armenia**  
*Martirosova, D., Inan, O. K., Meyer, M., Sinha, N.*, August 2017
- 119 **Natural Disaster Damage Indices Based on Remotely Sensed Data: An Application to Indonesia**  
*Skoufias, E., Strobl, E., Tveit, T. B.*, September 2017
- 120 **The Distributional Impact of Taxes and Social Spending in Croatia**  
*Inchauste Comboni, M. G., Rubil, I.*, October 2017
- 121 **Regressive or Progressive? The Effect of Tobacco Taxes in Ukraine**  
*Fuchs, A., Meneses, F.*, September 2017
- 122 **Fiscal Incidence in Belarus: A Commitment to Equity Analysis**  
*Bornukova, K., Shymanovich, G., Chubrik, A.*, October 2017

- 123 **Who escaped poverty and who was left behind? a non-parametric approach to explore welfare dynamics using cross-sections**  
*Lucchetti, L. R.,* October 2017
- 124 **Learning the impact of financial education when take-up is low**  
*Lara Ibarra, G., Mckenzie, D. J., Ruiz Ortega, C.,* November 2017
- 125 **Putting Your Money Where Your Mouth Is Geographic Targeting of World Bank Projects to the Bottom 40 Percent**  
*Öhler, H., Negre, M., Smets, L., Massari, R., Bogetić, Z.,* November 2017
- 126 **The impact of fiscal policy on inequality and poverty in Zambia**  
*De La Fuente, A., Rosales, M., Jellema, J. R.,* November 2017
- 127 **The Whys of Social Exclusion: Insights from Behavioral Economics**  
*Hoff, K., Walsh, J. S.,* December 2017
- 128 **Mission and the bottom line: performance incentives in a multi-goal organization**  
*Gine, X., Mansuri, G., Shrestha, S. A.,* December 2017
- 129 **Mobile Infrastructure and Rural Business Enterprises Evidence from Sim Registration Mandate in Niger**  
*Annan, F., Sanoh, A.,* December 2017
- 130 **Poverty from Space: Using High-Resolution Satellite Imagery for estimating Economic Well-Being**  
*Engstrom, R., Hersh, J., Newhouse, D.,* December 2017
- 131 **Winners Never Quit, Quitters Never Grow: Using Text Mining to measure Policy Volatility and its Link with Long-Term Growth in Latin America**  
*Calvo-Gonzalez, O., Eizmendi, A., Reyes, G.,* January 2018
- 132 **The Changing Way Governments talk about Poverty and Inequality: Evidence from two Centuries of Latin American Presidential Speeches**  
*Calvo-Gonzalez, O., Eizmendi, A., Reyes, G.,* January 2018
- 133 **Tobacco Price Elasticity and Tax Progressivity In Moldova**  
*Fuchs, A., Meneses, F.,* February 2018
- 134 **Informal Sector Heterogeneity and Income Inequality: Evidence from the Democratic Republic of Congo**  
*Adoho, F., Doumbia, D.,* February 2018
- 135 **South Caucasus in Motion: Economic and Social Mobility in Armenia, Azerbaijan and Georgia**  
*Tiwari, S., Cancho, C., Meyer, M.,* February 2018

- 136 **Human Capital Outflows: Selection into Migration from the Northern Triangle**  
*Del Carmen, G., Sousa, L., February 2018*
- 137 **Urban Transport Infrastructure and Household Welfare: Evidence from Colombia**  
*Pfutze, T., Rodriguez-Castelan, C., Valderrama-Gonzalez, D., February 2018*
- 138 **Hit and Run? Income Shocks and School Dropouts in Latin America**  
*Cerutti, P., Crivellaro, E., Reyes, G., Sousa, L., February 2018*
- 139 **Decentralization and Redistribution Irrigation Reform in Pakistan's Indus Basin**  
*Jacoby, H.G., Mansuri, G., Fatima, F., February 2018*
- 140 **Governing the Commons? Water and Power in Pakistan's Indus Basin**  
*Jacoby, H.G., Mansuri, G., February 2018*
- 141 **The State of Jobs in Post-Conflict Areas of Sri Lanka**  
*Newhouse, D., Silwal, A. R., February 2018*
- 142 **"If it's already tough, imagine for me..." A Qualitative Perspective on Youth Out of School and Out of Work in Brazil**  
*Machado, A.L., Muller, M., March 2018*
- 143 **The reallocation of district-level spending and natural disasters: evidence from Indonesia**  
*Skoufias, E., Strobl, E., Tveit, T. B., March 2018*
- 144 **Gender Differences in Poverty and Household Composition through the Life-cycle A Global Perspective**  
*Munoz, A. M., Buitrago, P., Leroy de la Briere, B., Newhouse, D., Rubiano, E., Scott, K., Suarez-Becerra, P., March 2018*
- 145 **Analysis of the Mismatch between Tanzania Household Budget Survey and National Panel Survey Data in Poverty & Inequality Levels and Trends**  
*Fuchs, A., Del Carmen, G., Kechia Mukong, A., March 2018*
- 146 **Long-Run Impacts of Increasing Tobacco Taxes: Evidence from South Africa**  
*Hassine Belghith, N.B., Lopera, M. A., Etang Ndip, A., Karamba, W., March 2018*
- 147 **The Distributional Impact of the Fiscal System in Albania**  
*Davalos, M., Robayo-Abril, M., Shehaj, E., Gjika, A., March 2018*
- 148 **Analysis Growth, Safety Nets and Poverty: Assessing Progress in Ethiopia from 1996 to 2011**  
*Vargas Hill, R., Tsehaye, E., March 2018*
- 149 **The Economics of the Gender Wage Gap in Armenia**  
*Rodriguez-Chamussy, L., Sinha, N., Atencio, A., April 2018*

- 150 **Do Demographics Matter for African Child Poverty?**  
*Batana, Y., Cockburn, J.,* May 2018
- 151 **Household Expenditure and Poverty Measures in 60 Minutes: A New Approach with Results from Mogadishu**  
*Pape, U., Mistiaen, J.,* May 2018
- 152 **Inequality of Opportunity in South Caucasus**  
*Fuchs, A., Tiwari, S., Rizal Shidiq, A.,* May 2018
- 153 **Welfare Dynamics in Colombia: Results from Synthetic Panels**  
*Balcazar, C.F., Dang, H-A., Malasquez, E., Olivieri, S., Pico, J.,* May 2018
- 154 **Social Protection in Niger: What Have Shocks and Time Got to Say?**  
*Annan, F., Sanoh, A.,* May 2018
- 155 **Quantifying the impacts of capturing territory from the government in the Republic of Yemen**  
*Tandon, S.,* May 2018
- 156 **The Road to Recovery: The Role of Poverty in the Exposure, Vulnerability and Resilience to Floods in Accra**  
*Erman, A., Motte, E., Goyal, R., Asare, A., Takamatsu, S., Chen, X., Malgioglio, S., Skinner, A., Yoshida, N., Hallegatte, S.,* June 2018
- 157 **Small Area Estimation of Poverty under Structural Change**  
*Lange, S., Pape, U., Pütz, P.,* June 2018
- 158 **The Devil Is in the Details; Growth, Polarization, and Poverty Reduction in Africa in the Past Two Decades**  
*F. Clementi F., Fabiani, M., Molini, V.,* June 2018
- 159 **Impact of Conflict on Adolescent Girls in South Sudan**  
*Pape, U., Phipps, V.,* July 2018
- 160 **Urbanization in Kazakhstan; Desirable Cities, Unaffordable Housing, and the Missing Rental Market**  
*Seitz, W.,* July 2018
- 161 **SInequality in Earnings and Adverse Shocks in Early Adulthood**  
*Tien, B., Adoho, F.,* August 2018
- 162 **Eliciting Accurate Responses to Consumption Questions among IDPs in South Sudan Using “Honesty Primes”**  
*Kaplan, L., Pale, U., Walsh, J.,* August 2018

163 **What Can We (Machine) Learn about Welfare Dynamics from Cross-Sectional Data?**

*Lucchetti, L.*, August 2018

164 **Infrastructure, Value Chains, and Economic Upgrades**

*Luo, X., Xu, X.*, August 2018

165 **The Distributional Effects of Tobacco Taxation; The Evidence of White and Clove Cigarettes in Indonesia**

*Fuchs, A., Del Carmen, G.*, August 2018

For the latest and sortable directory,  
available on the Poverty & Equity GP intranet site. <http://POVERTY>

[WWW.WORLDBANK.ORG/POVERTY](http://WWW.WORLDBANK.ORG/POVERTY)