

Welfare Dynamics in Colombia

Results from Synthetic Panels

Carlos Felipe Balcazar

Hai-Anh Dang

Eduardo Malasquez

Sergio Olivieri

Julieth Pico



WORLD BANK GROUP

Poverty and Equity Global Practice

&

Development Data Group

May 2018

Abstract

This study explores the short-run transitions between poverty, vulnerability, and middle class, using synthetic panels constructed from multiple rounds of Colombia's Integrated Household Survey (in Spanish Gran Encuesta Integrada de Hogares). The paper reports results from two approaches to define a vulnerability line: the first one employs a nonparametric and parsimonious model, while the second utilizes a fully parametric regression model with covariates. The estimation results suggest a range of between \$8 to \$13 per

day per person in 2005 purchasing power parity dollars as the vulnerability line. Using an average daily vulnerability line of \$10 per day per person, subsequent estimates on welfare dynamics suggest that, during the past decade, 20 percent of the Colombian population experienced downward mobility, and 24 percent experienced upward mobility. Furthermore, upward mobility increases with higher education levels and is lower for female-headed households.

This paper is a product of the Poverty and Equity Global Practice and the Development Data Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at solivieri@worldbank.org.

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

Welfare Dynamics in Colombia: Results from Synthetic Panels

Carlos Felipe Balcazar, Hai-Anh Dang,
Eduardo Malasquez, Sergio Olivieri and Julieth Pico*

JEL Codes: C14, D31, I32

Keywords: welfare dynamics, poverty, vulnerability, middle class, panel data

* Balcazar is a PhD student, Wilf Family Department of Politics, New York University. Dang is an Economist in the Survey Unit, Development Data Group, World Bank. Malasquez is an Economist in the Poverty and Equity Global Practice, World Bank. Olivieri is Senior Economist in the Poverty and Equity Global Practice, World Bank. Pico is a Consultant in the Poverty and Equity Global Practice, World Bank. We would like to thank Leonardo Lucchetti and Mariana Viollaz for comments on earlier versions of the paper. We would also like to thank the UK Department of International Development for funding assistance through its Strategic Research Program (SRP) program.

1. Introduction

Colombia's recent record of solid economic growth led to significant reductions in poverty and improvements in social indicators from 2002 to 2016. Moreover, during this period, the extreme poverty rate more than halved, falling from 17.7 percent to 8.5 percent; while moderate poverty fell from 49.7 percent to 28.0 percent, as shown in Figure 1. Consequently, the number of poor in Colombia declined from about 20 million in 2002 to approximately 13.2 million in 2016. In addition, during the 2008-2016 period the Gini index declined from 56.7 to 51.7. However, Colombia continues to be one of the most unequal countries in the region (see the results on the Gini index in Figure 1) with higher inequality than the average of the region (the Gini index for the Latin America and the Caribbean region was 51.4 in 2014 versus 53.8 in Colombia for the same year) and neighboring countries such as Ecuador (45.4 in 2014), Panama (50.7 in 2014) and Peru (44.1 in 2014).

Despite showing a downward trend in poverty from 2002 to 2015, monetary poverty did not show statistically significant changes from 2015 to 2016 (poverty increased 0.2 percentage points), while extreme poverty increased slightly during the same period (0.6 percentage points). This behavior is explained by the reduction in the growth rate experienced by Colombia, after the plunge in oil prices in mid-2014, followed by a devaluation of the Colombian peso in 2015 and high inflation in 2016. This raises concerns regarding the vulnerability of households to falling into poverty as well as on the evolution of the middle class during difficult economic environments. Thus, additional measures of welfare dynamics are necessary to obtain a richer picture of the evolution of movements across the distribution.

Ideally, a researcher would like to use a longitudinal survey or panel data to analyze welfare dynamics or income mobility. However, in many developing countries panel data sets are not readily available, span few periods, or suffer from “non-random” attrition issues, hindering the capacity of researchers to study elements such as the factors that help households escape or remain in poverty (Dang and Lanjouw, 2013; Bourguignon and Moreno, 2015). To overcome the absence of panel data or longitudinal surveys, authors such as Deaton (1985), Deaton and Paxson (1994) and Pencavel (2007) have proposed methodologies to construct pseudo-panels by following similar age cohorts across multiple cross-section surveys. Nevertheless, as argued by Dang *et al.* (2014), these methodologies typically rely on having several rounds of cross-section surveys, but do not allow to analyze mobility at a more disaggregated level than the cohort. In addition, Fields and Viollaz (2013) argue that pseudo-panel methodologies might not perform well in predicting income mobility in some cases.¹

¹ Fields and Viollaz (2013) analyze the performance of two pseudo-panels methods to predict income mobility only for Chile.

Dang *et al.* (2014) propose both a parametric and a non-parametric approach to construct synthetic panels and estimate an upper-bound (assuming zero correlation between error terms) and a lower-bound (assuming perfect positive correlation between error terms) for the transitions using two rounds of cross sections. In addition, in Dang and Lanjouw (2013, 2016) the authors extend this method further and calculate point estimates of poverty mobility based on the synthetic-panels (relying on the key assumption that the residual terms of the income equations in two periods are distributed according to a bi-normal distribution).² In addition to the analysis of the transitions in and out of poverty using synthetic-panels, Dang and Lanjouw (2017) estimate vulnerability lines to analyze welfare dynamics in India (note that the authors also use true panel data from Vietnam and the United States to illustrate their method to assess vulnerability), connecting the specification of the vulnerability line to the risk of falling into poverty.

The purpose of this paper is twofold. First, define the vulnerability lines, and second analyze the short-run transitions between poverty, vulnerability and middle class in Colombia, relying on an application of the synthetic panel methodology proposed by Dang and Lanjouw (2013, 2016) and Dang *et al.* (2014) using as few as two rounds of cross-section surveys and relatively parsimonious modeling assumptions. The approach models income or consumption using only time-invariant covariates common to two (or more) surveys. The procedure assumes that the underlying populations being sampled in both rounds of the survey are the same, such that the time-invariant household characteristics in one round of the survey would be the same in the following round. The main implication of this assumption is that households interviewed in the second period of the survey with characteristics similar to those of households interviewed in the first period would have achieved similar levels of income or consumption in the initial period and vice versa, providing the linkage between household income or consumption in both periods.

One of the cross-cutting strategies envisioned to achieve the strategic development objectives of the country per the National Development Plan of Colombia for the 2014-2018 period³ is promoting social mobility. In this context, the National Development Plan highlights the need to deepen the analysis of inequality and welfare dynamics in Colombia. Information on the dynamics of income allows to identify the characteristics associated with economic mobility, as well as identify the segment of the population more vulnerable to return to poverty. For the Latin America region, the World Bank defines the middle class as those individuals with a per capita income from US\$10 to US\$50 per day in purchasing power parity (PPP) at 2005 international prices, values estimated using the vulnerability to poverty approach proposed by Lopez-Calva and Ortiz-Juarez (2014). This methodology assumes that households with incomes above the poverty line but with a probability of falling into poverty higher than 10 percent should

² In addition, a recent paper by Bourguignon and Moreno (2015) discusses using two cross sections to construct synthetic panels relying on pseudo-panel, matching and calibration techniques.

³ See the National Development Plan at: <https://colaboracion.dnp.gov.co/CDT/PND/PND%202014-2018%20Tomo%201%20internet.pdf>

be classified as vulnerable. Notice, however, that the methodology proposed by Lopez-Calva and Ortiz-Juarez (2014) requires at least two waves of longitudinal information or panel data, which are not currently available in Colombia.

Colombia has two publicly available surveys designed to follow households across multiple periods: (i) the *Encuesta Longitudinal de Protección Social* (ELPS)⁴ prepared by the Colombian statistics office (DANE), and (ii) the *Encuesta Longitudinal Colombiana* (ELCA)⁵ elaborated by the *Universidad de los Andes*. However, in practice there is only one publicly available wave of the ELPS (making it a cross-section for concrete purposes) and, while there are two waves available of the ELCA, the survey does not properly capture incomes (the official welfare measure used for estimating extreme and moderate monetary poverty), raising concerns on the comparability of poverty estimates obtained using the ELCA and the GEIH. Therefore, the use of alternative strategies, such as the one proposed in this paper, is required to assess the dynamics of households living in poverty, vulnerability or the middle class.

Using cross-section information from multiple rounds of the *Gran Encuesta Integrada de Hogares* (GEIH), this paper constructs a synthetic-panel for Colombia, and based on the methodology proposed by Dang et al. (2016), this paper estimates the vulnerability lines relevant for Colombia during the period 2008-2016, and calculates the transitions between poverty, vulnerability, and middle-class. Additionally, the paper shows several sensitivity analyses to better inform the crucial decision of choosing a vulnerability line. Results suggest a US\$ 10 dollar-a-day in 2005 PPP (i.e. US\$ 13.2 dollar-a-day in 2011 PPP) as the vulnerability threshold for Colombia. The “monetary welfare” dynamics suggest that roughly 56 percent of the Colombian population remain in the same income categories, 20 percent experience downward mobility and the remaining 24 percent experienced upward mobility. Furthermore, we observe that the rate of escaping poverty and vulnerability into the middle class, and the rate of escaping poverty into vulnerability increases with the levels of education, and it is lower for female household heads than their counterpart male household heads.

The rest of this study is organized as follows. The next section discusses the methodology proposed by Dang *et al.* (2014) to construct synthetic panels based on cross-section surveys to analyze poverty dynamics⁶ with an application for Colombia. The third section shows the main characteristics of the data available for Colombia during the 2008-2016 period and, more importantly, of the relevant sample used for this study as well as how the window width of the synthetic panel was defined. Section 4 presents a sensitivity analysis for the vulnerability lines in Colombia using different base years and alternative

⁴ <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/encuesta-longitudinal-de-proteccion-social-elps>

⁵ <https://encuestalongitudinal.uniandes.edu.co/en/>

⁶ In this paper, the term poverty dynamics is associate to joint probabilities, and not to conditional probabilities. For example, the joint probability of being poor in t and being poor in $t+1$.

methods to estimate the vulnerability lines. The fifth section shows preliminary results and discuss briefly welfare dynamics across multiple potential states (i.e. poor, vulnerable and middle class). The last section presents some final remarks.

2. Methodology

A proper study of welfare dynamics typically entails a demanding minimum set of data requirements. It is necessary to follow the same observation (household or individual) for at least two—or preferably, multiple—periods. However, panel data or longitudinal data sets are hard to come by, especially in developing countries, while “snap-shots” of welfare captured in cross-section surveys are far more common (Dang and Lanjouw, 2013). This paper proposes to rely on a synthetic panel approach to provide point estimates of the income mobility in Colombia using as few as two rounds of cross section surveys. Moreover, this study extends the typical analysis of transition in and out of poverty to analyze a more general setup of household movements across different income groups (poor, vulnerable and middle class).

The approach is intended to overcome the lack of available panel data by constructing a “synthetic panel” using only time-invariant individual and household characteristics from multiple rounds of the *Gran Encuesta Integrada de Hogares* (GEIH) of Colombia and exploiting this information to estimate the vulnerability lines necessary for the analysis of welfare dynamics. First, the following section explains the estimation of the vulnerability line and then presents an overview of the methodology used to study the transitions across poverty, vulnerability and middle class.

2.1. Vulnerability lines

On occasions researchers or policy makers are interested in studying more than the transitions in and out of poverty. In Colombia, where a large share of the population has escaped poverty during the last decade—despite the country was recently exposed to downside risks from volatile commodity prices—there is an increasing interest in identifying the dynamics into the condition of vulnerability (i.e. population out of poverty but at risk of falling back into poverty, hence being *vulnerable*) and the middle class. In this context, the discussion on the estimation of a vulnerable group in Colombia is relevant from both a technical and a public policy perspective.

Since true panel data are not available in Colombia, this paper proposes to rely on the approach by Dang and Lanjouw (2017) to estimate vulnerability lines, using as few as two rounds of cross sections and moderate assumptions. Dang and Lanjouw (2017) define the vulnerability line V_1 such that a specified proportion of the population with a consumption level above this line in period 1 will fall below the poverty line Z in period 2. This proportion is referred to as the “insecurity index”, P^1 , since the population with

income levels above the vulnerability line could be regarded as “secure”. Given a value for the “insecurity index” P^1 , then V_1 satisfies:

$$P^1 = P(y_2 \leq Z | y_1 > V_1)$$

In addition, the definition of the insecurity index could be linked to a notion of “secure” population, which has incomes above the poverty line but still below the vulnerability line in period 1. The likelihood among this population of falling into poverty in period 2 is the “vulnerability index” P^2 and satisfies:

$$P^2 = P(y_2 \leq Z | Z < y_1 < V_1)$$

Both the “insecurity index” and “vulnerability index” provide operational measures for households’ vulnerability to poverty, but while the vulnerability index focuses in the population in the middle of the income or consumption distribution, the insecurity index focuses on households located in the top of such distribution. Figure 2, taken from Dang and Lanjouw (2017), shows the differences between the insecurity index and the vulnerability index and how they relate.

2.2. Overview of the framework

As an introduction to the synthetic panel methodology proposed by Dang and Lanjouw (2016), this section summarizes the framework used to construct synthetic panel data from two rounds of cross sectional data. Assume there are two rounds of cross sectional surveys such that y_{ij} is the corresponding income for individual $i = \{1, 2, \dots, N\}$ in survey round $j = \{1, 2\}$, with sample size N . Now, let $x_{i,j}$ be a vector of household characteristics. These variables can be either time-invariant (e.g., gender, ethnicity, language, place of birth, etc.), variables that can be easily recalled for round 1 in round 2 (e.g., information about household heads’ age, education, etc.), or retrospective regressors.

Using these variables, the linear projection of household’s “ i ” income (or consumption) $y_{i,j}$ on household characteristics for each survey round “ j ” is given by:

$$(1) \quad y_{i,j} = \beta_j' x_{i,j} + \varepsilon_{i,j}$$

If we are only interested in studying poverty dynamics, (using both incomes $y_{i,j}$ and the poverty line z are expressed in real terms), we are interested in knowing such quantities as

$$(2) \quad P(y_{i1} < z_1 \text{ and } y_{i2} > z_2)$$

which represents the percentage of households that are poor in the first period but non-poor in the second period (see appendix A, for a more detailed explanation). Nevertheless, when we are interested in studying the dynamics between poverty and vulnerability, we are interested in such quantities as

$$(3) \quad P(y_{i1} < z_1 \text{ and } z_2 < y_{i2} < v_2)$$

which represents the percentage of poor households in the first period that move into the vulnerable category in the second period. There are in total nine combinations of income categories when two periods are considered (see appendix B, for a more detailed explanation).

In the absence of true panel data, we need to use synthetic panels to study mobility, making two standard assumptions. Following Dang and Lanjouw (2013), the first assumption is that the underlying populations being sampled in survey rounds 1 and 2 are the same in terms of the time-invariant household characteristics. The second one is that ε_{i1} and ε_{i2} have a bivariate normal distribution⁷ with the (partial) correlation coefficient ρ and standard deviations $\sigma_{\varepsilon 1}$ and $\sigma_{\varepsilon 2}$ respectively. If ρ is known, Dang and Lanjouw (2013) propose to estimate quantity (3) by

$$(4) \quad P(y_{i1} < z_1 \text{ and } z_2 < y_{i2} < v_2) = \Phi_2\left(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{\varepsilon 1}}, \frac{v_2 - \beta_2' x_{i2}}{\sigma_{\varepsilon 1}}, \rho\right) - \Phi_2\left(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{\varepsilon 1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{\varepsilon 1}}, \rho\right)$$

where $\Phi_2(\cdot)$ stands for the bivariate normal cumulative distribution function (cdf). A key element in the analysis of income mobility is the estimation of the correlation coefficient. Since ρ is usually unknown in most contexts, it is possible to obtain an approximation based on asymptotic theory following the approach proposed by Dang and Lanjouw (2016). The procedure requires aggregating all the variables to the cohort level, where cohorts are formed by a different combination of all the values of the time-invariant characteristics (including age, gender, and education):

$$(5) \quad y_{cj} = \beta_j' x_{cj} + \varepsilon_{cj}$$

Then, the partial correlation coefficient can be estimated as follows,

$$(6) \quad \rho_{\varepsilon c1 \varepsilon c2} = \frac{cov(\varepsilon_{c1}, \varepsilon_{c2})}{\sqrt{var(\varepsilon_{c1}) var(\varepsilon_{c2})}}$$

⁷ Formal multivariate normality tests (Doornik and Hansen, 2008) reject the assumption of a normal distribution (univariate or bivariate) for the error terms in Colombia. But as noted in Dang et al. (2014) and Dang and Lanjouw (2013), since these tests are rather demanding, the method may still perform reasonably well even where the tests are not strictly satisfied.

The estimates of β_j correspond to the linear projection on household income or consumption on household characteristics aggregated at the year of birth level for survey round $j = 1, 2$.

3. Data

This section of the paper discusses the characteristics of the main source of information used to construct the synthetic panel for Colombia during the 2008-2016 period — the *Gran Encuesta Integrada de Hogares* (GEIH) — and is divided in three subsections. The first part shows the main characteristics of the GEIH, the cross-sectional national household survey that provides the data to build each wave of the synthetic panel. After analyzing the main characteristics of the population in Colombia during the period 2008-2016, the second part describes the relevant sample: all households included in the surveys from 2008 to 2016 whose heads were born between 1948 and 1973. The rationality behind this selection is that household heads in this cohort who were interviewed in 2008 (the earlier wave of the GEIH analyzed) are expected to have completed their education—at least 25 years old—and still be part of the labor force— younger than 60 years. As mentioned on Lucchetti (2017), the selection of these household heads also avoids life cycle events that may invalidate the time-invariant assumption. Once this cohort is fixed, the methodology suggests following the same cohort of individuals across time. This section ends with a discussion on the selection of the optimal window of analysis to build the synthetic panel, or the distance in years between two cross-sections of the survey. This paper argues that such window for Colombia should not be longer than two years, since for longer gaps it is likely that the characteristics of the households would change significantly, thus violating the assumption of time invariability of the characteristics associated with the income generation function.

3.1. Main source of information: The GEIH

The GEIH is a nationally representative survey administered by the National Administrative Department of Statistics (DANE,⁸ for its acronym in Spanish) that captures information on the employment conditions of individuals and is the main source of information to estimate monetary poverty in Colombia. Currently, the GEIH interviews approximately 240,000 households every year, being the largest survey at the national level in Colombia. In addition to the general characteristics of the population such as gender, age, marital status and educational level, the GEIH asks about different sources of income. The survey classifies the workforce in one of three groups: occupied, unemployed or inactive. Using the information in the GEIH it is possible to estimate the main indicators of the Colombian labor market, such as the global participation rate, the occupancy rate, and the unemployment rate.

⁸ *Departamento Administrativo Nacional de Estadísticas*—DANE.

The GEIH has national coverage with the following levels of temporal and geographical disaggregation: (i) Monthly: for a group of 13 large capital cities and their metropolitan areas, 11 intermediate capital cities and for the national total. (ii) Quarterly: by capital city (large and intermediate) with its corresponding metropolitan area and for the total of the country by zone (head of the department, populated centers and dispersed rural areas). (iii) Biannual: by capital city, by large regions (Atlantic, Eastern, Central, Pacific and Bogota) and for headlands and populated centers and rural dispersed and for the national total per zone (headers and populated centers and rural dispersed). (iv) Yearly: by capital city with its metropolitan area, by large regions and area (headlands and populated centers and dispersed rural) and by departments.

3.2. The characteristics of the synthetic panel

Following Dang and Lanjouw (2013) and Dang et al. (2014) it is important to verify that the distributions of the time-invariant variables for the two survey rounds are similar across different periods, since the proposed approach relies on the assumption that both surveys represent the same population and that income can be modeled based on such time invariant characteristics. **Table 1** shows the estimated poverty rates for Colombia during the 2008-2016 period using both samples: the full sample of observations available in the GEIH and the restricted sample which includes only household heads whose age goes from 25 to 60 years old in 2008 (i.e. household heads who were born between 1948 and 1973). The final sample includes 1,479,549 households with an average of 184,944 households per survey. The estimates in **Table 1** suggest that the restricted sample reproduces very closely the moderate official poverty estimates for Colombia during the period of analysis. The average difference in poverty rates associated with the full and the restricted sample is 0.40 percent points from 2008 to 2016. This suggests that our estimation sample reflects adequately the Colombian population's poverty rates as measured in the unrestricted cross sections.

The second step is to assess whether the GEIH rounds are strictly comparable. Our findings suggest that the survey rounds do not appear to suffer from serious comparability issues, especially the potentially time invariant variables for the income model. We focus on household heads, which represent 28 percent of the population, to implement the synthetic-panel methodology. In addition, this paper uses the survey design of the GEIH to improve the precision of the estimates presented in this section. The same cohort of individuals is followed across time to implement the methodology proposed by Dang and Lanjouw (2016). The variables chosen to construct the synthetic panels are the following: birth year (cohort group), gender,

and education attainment (level). There are only 0.04 percent missing values related to the previous variables.⁹

Given that one of the time invariant characteristics chosen for the analysis is the level of education of the household head, we restrict the sample to individuals from 25 to 68 years of age.¹⁰ This decision is made to avoid truncation in the variable of educational attainment, and to guarantee representativeness of household heads (the implicit assumption being that by 25 years old the average Colombian should have completed his or her education). In addition, restricting the head of the household head's age to a specific range is a standard procedure to keep the household composition stable over different periods.

Table 2 below shows the composition of the sample, based on different observable characteristics, for each survey from 2008 to 2016. The table suggests that the share of heads from older cohorts (the ones born in the 1948-1953 and 1954-1963 period) tends to fall—especially for the heads born between 1954 and 1963, falling from 40 to 36 percent—while the share of heads belonging to younger cohorts—from 1964 to 1973—increases 6 percentage points during the period of analysis. In addition, taking 2008 as the reference period, the share of female household heads increases over time and these changes are significant. Moreover, the population also tends to become more educated over time, although the share of household heads with no formal education remained relatively constant across all the periods of analysis.

3.3. Defining the relevant window of analysis

To define the time interval between two cross sections in which the assumption of invariability holds, the different characteristics of the household heads are formally tested (see **Table C.1. - Table C.6.** in the Appendix C). This procedure implements a t-test of the means for each of the time invariant characteristics at different periods to determine if they are not statistically different. The results suggest that the assumption of time invariability is less plausible when comparing pairs of surveys more than two years apart from each other. For instance, **Table C.1.** suggests the gender composition of household heads in the synthetic panel is one of the characteristics that changes faster across time than any other characteristic. The mean test estimates in **Table C.1** indicate that these characteristics change significantly from one year to another, undermining the assumption of time invariability. In contrast, other characteristics, such as the share of heads of household with no formal education, are relatively stable across the whole period of analysis, as shown by **Table C.2.** The assumption of time invariability seems to be consistent with the empirical findings but will not necessarily hold for other education levels. For instance, for the share of

⁹ As part of the analysis, we identify income per capita outliers using the Blocked adaptive computationally efficient outlier nominators (BACON). After applying this method, this paper finds that 0.1 percent of the sample observations are classified as outliers and they are present mostly in 2008 (that is approximately 1 percent of the sample in 2008).

¹⁰ As the age range should be kept fixed over time for all the different cohorts (i.e. adjusting for the year difference between the survey rounds), we should use the age range 25-60 for 2008, 27-62 for 2010, 29-64 for 2012 and so on.

household heads only with primary education (**Table C.3**), differences are statistically significant for most cases when comparing pairs of surveys with more than one year of separation. In sum, the results show the differences in education levels of the heads of households across different years are significant, suggesting caution and providing evidence to be conservative when relying in the assumption of time invariability of the characteristics across periods. Strictly speaking, the results of the means tests suggest that for Colombia the time interval between the cross-section surveys should not be more than two years apart from each other.

4. Vulnerability lines

Identifying the vulnerable group usually relies on estimating “appropriate” lines that allow us to classify the population in different categories, similarly to the definition of poverty. However, in contrast with poverty, where typically one threshold (i.e. poverty line) is enough to split the population in poor and non-poor, in the case of vulnerability two lines might be necessary. First, we need to define the cutoff point that represents the lower bound of the vulnerable group, which in practice usually coincides with the poverty line, meaning that people or households who graduate from poverty (i.e. achieve incomes above the poverty line) do not immediately become part of the middle class but instead remain in a state of vulnerability. Second, probably the most relevant line is the upper bound of the vulnerable group that we call “vulnerability line”, and typically represents the lower bound of the middle class.

Once poverty and vulnerability lines are defined, individuals or households could be classified either as poor when their incomes are below the poverty line, or as vulnerable when their incomes are above the poverty line but below the vulnerability line, or as middle class when their incomes are above the vulnerability line (and implicitly above the poverty line, since the value of the vulnerability line is higher than the poverty line). Notice, the economic literature (Atkinson and Brandolini, 2013; Lopez-Calva and Ortiz-Juarez, 2014) usually considers the upper bound line of the vulnerable group as the lower bound of the middle class. In addition, any proposal to empirically estimate a vulnerable group requires accepting the implicit assumption that it is possible not only to formulate a relevant concept for class but also to identify these categories from empirical methods (Lopez-Calva and Ortiz-Juarez, 2014).

Although there is currently no consensus in the literature on the best methodology to estimate the vulnerability line, the Government of Colombia has been exploring the use of the economic security approach, based on the criterion of vulnerability to poverty, to identify the upper limit of the vulnerable group or the lower bound of the middle class based on the Lopez-Calva and Ortiz-Juarez approach (see

Pavon and Perez, 2016).¹¹ Even though these results based on the ELCA were informative, they raised several concerns since the poverty figures were not only different from official estimates but also reflected a different poverty line (i.e. international poverty lines were used in this exercise). However, the main requirement the Lopez-Calva and Ortiz-Juarez (2014) methodology has is the existence of longitudinal information or panel data, which are not currently available in Colombia.

It is important to note that, for the Dang and Lanjouw (2017) approach, there is no close solution for V_1 that can be obtained from the equations of the “insecurity” and “vulnerability” indexes. However, given household income in both periods, the poverty line z , and some pre-determined value for either the insecurity or vulnerability index, it is possible to empirically solve for the vulnerability line V_1 . The construction of the vulnerability lines can be approached as a two-step process: the first is to identify the appropriate poverty lines (which are usually given; for example, the international poverty line); the second step is to iterate upward from the given poverty line, until we reach a value of the vulnerability line that provides the specified vulnerability index.

This method produces a set of vulnerability lines for a set of vulnerability indexes, posing the challenge for policy makers or society to choose a number from that set. The rule of thumb used under this approach for developing countries is to set a vulnerability index between 15 and 30 percent (or as desired by the social development objectives). This identification difficulty is not particular to this approach. For instance, under the Lopez-Calva and Ortiz-Juarez (2014) method, a figure is defined according to a “certain” risk level of falling back into poverty, which is usually 10 percent; some other rules have been used to define the middle class.¹² There is no clear consensus in the literature on how to objectively identify this line; instead it looks more like an arbitrary decision either to choose the likelihood level to fall into poverty or a number from a set of vulnerability indexes. Two exercises are presented to better inform this crucial decision: the first one consists on a sensitivity analysis of vulnerability lines when the base year is changed; and the second one is inspired by the method by Hertova, Lopez-Calva and Ortiz-Juarez (2010).

4.1. Estimation of the vulnerability line

Poverty lines are sensitive to the base year used for their estimation. These lines are drawn from particular welfare distributions that change over time given the underlying development process in which many

¹¹ The vulnerability to poverty approach can be divided in three steps. The first stage focuses on identifying the characteristics associated with the transition to and from the poverty condition. The second stage seeks to model the probability of falling into poverty, based on a series of observable variables, using a logistic model. The third step uses the variables that explain the probability of falling into poverty to predict the expected income associated with each level of probability. The third stage allows to identify the level of income associated with a 10 percent probability of falling into poverty (per the stylized facts reported by Cruces et al., 2011).

¹² Lawrence (1984), Blackburn & Bloom (1987), Horrigan & Haugen (1988), Kusters & Ross (1989), Birdsall, Graham & Pettinato (2000) and D'Ambrosio et al (2002), Atkinson, A. and Brandolini, A. (2011) to quote a few.

different forces or factors are involved, such as consumption patterns change over time. Similarly, vulnerability lines are expected to change over time. Even though, efforts have been made to maintain the characteristics of cohorts fixed over time, in the Colombia case the maximum interval window time was two years for constructing the synthetic panels, some variation is expected when the base year is changed on vulnerability lines over a set of vulnerability indexes.

The sensitivity exercise starts by using the 2008-2010 synthetic panel and estimates the set of vulnerability lines for a range of vulnerability indexes between 22 and 35 percent. **Table 3** shows the main results of the estimation of alternative vulnerability lines for different levels of the vulnerability index for Colombia. The second column shows the share of the vulnerable population in the first period that becomes poor in the second period, or vulnerability index, associated with the vulnerability line expressed in COP in column (3) and in \$PPP 2011 in column (4). Column (5) shows the percentage increase in the poverty line corresponding to the value of the vulnerability line, and column (6) shows the percentage of people with consumption below the vulnerability line and above the poverty line in 2008 and 2010.

Repeating the exercise for other pairs of years (i.e. 2010-2012, 2012-2014, 2014-2016) and the same vulnerability indexes, four sets of vulnerability lines are found. **Figure 3** presents the combination of vulnerability lines and vulnerability indexes when different couples of years are used. From simple inspection, there is certain convergence of the vulnerability lines over time as the vulnerability index increases. This event seems to happen when the vulnerability index is within 30-32 percent which represents an increment of around two times the poverty line.

4.2. An alternative approach to define the vulnerability line

This section presents the results of implementing the Hertova, Lopez-Calva and Ortiz-Juarez (2010) approach to estimate the vulnerable households. The idea is to shed some light on the robustness of the results found so far. However, it is relevant to notice that the empirical method was adapted to find a vulnerability line.

In recent years, authors like Atkinson and Brandolini (2013) and Lopez-Calva and Ortiz-Juarez (2014) have proposed to study vulnerability anchoring the concept to the risk of falling into poverty. In line with Ravallion (2010), these authors suggest that although during recent decades, the population in developing countries seems to be escaping poverty, some of the households moving beyond the poverty threshold are still highly vulnerable and only marginally better off than their “poor” counterparts. In this context, the key element to define the middle class would be how safe are the income based middle-class citizens from falling back into poverty.

In the case of Colombia, one of the limitations to apply the methodology proposed by Lopez-Calva and Ortiz-Juarez (2014) is that there is no publicly available longitudinal survey that allows to map the

different probabilities of falling into poverty with a specific level of income or consumption.¹³ Therefore, this paper takes the approach proposed by Hertova, Lopez-Calva and Ortiz-Juarez (2010). The authors use cross-section data surveys to determine the vulnerable population; we adapt this methodology to find the amount of comparable income associated with a 10 percent risk of falling into poverty (as suggested in Lopez-Calva and Ortiz-Juarez (2014)).

Note that, in contrast with the regressions in Section 3, the set of control variables or observable characteristics is not limited to time invariant characteristics. **Table 4** shows the estimated vulnerability lines for four alternative specifications:

- Specification 1: only includes variables associated with time invariable characteristics of the head of the household (i.e., education level, year of birth and gender).
- Specification 2: adds a set of variables associated with labor market outcomes (i.e., employment status, sector of the economy, type of employment).
- Specification 3: adds characteristics of the household such as household size and access to basic services.
- Specification 4: adds to specification 3, additional controls for exposure to shocks such as losing a job.

The results suggest that the vulnerability line is sensitive to the model specification. For instance, models controlling for a larger set of variables tend to produce lower income estimates associated with a 10 percent risk of falling into poverty.¹⁴ A plausible explanation is that models including more control variables do a better job capturing the elements associated with the probability of falling back into poverty, thus given that the model already controls for such elements the level of income associated with a particular risk of transition into poverty is lower.

4.3. Identifying the vulnerability line

Even though there is no objective method to identify the vulnerability line, the final choice of the vulnerability line seeks to be informed by the results from previous exercises. From the sensitivity analysis, a convergence pattern of the vulnerability index seems to happen within the 30-32 percent across time. This

¹³ Recent advances with synthetic panel techniques such as that of Bourguignon and Moreno (2015) may be applied to address this issue. Other alternative approaches have been proposed that aim to construct some measure of income mobility based on averaging the error terms of the household consumption model in some way (see, e.g., Stampini et al. (2016) and Lucchetti (2017)), but we would like to caution against such approaches since these studies do not offer an underlying theory that supports doing so.

¹⁴ We find the same patterns as Lopez-Calva and Ortiz-Juarez (2014), each time the authors add control variables the vulnerability line become smaller. For example, for Chile the vulnerability line when they control just for education, sex and age of heads is \$9.6 PPP 2005, and when they include locational effects, marital status and measures of changes the vulnerability line is \$8.5 PPP 2005.

represents a vulnerability line within the US\$ 8 and US\$ 11 dollar-a-day in 2005 PPP interval (i.e. US\$ 10 to US\$ 14.9 dollar-a-day in 2011 PPP). The second exercise shows that the vulnerability line would belong to an interval between US\$ 8.7 and US\$ 13.6 dollar-a-day in 2005 PPP (i.e. US\$ 11.5 and US\$ 17.8 dollar-a-day in 2011 PPP) for a 10 percent probability of falling into poverty considering all models and time intervals. Moreover, previous research on the value of the vulnerable group and the middle class finds that the value of the vulnerability line is around US\$ 9.5 to US\$10 (i.e. US\$ 12.6 and US\$13.2 dollar-a-day in 2011 PPP) (see Pabon and Perez, 2016; Lopez-Calva and Ortiz-Juarez, 2014).

The interval values for the vulnerability line estimated in previous exercises and research overlap within a range of US\$ 8 and US\$ 14 dollar-a-day 2005 PPP. However, this is not enough for identifying the vulnerability line. To do so, we first restrict the findings only to the 2008-10 synthetic panel, because it is the closest period when the official poverty lines were estimated. The range still is around US\$ 8.4 and US\$ 13.4 dollar-a-day 2005 PPP based on the sensitivity analysis and the adapted Lopez-Calva Method. The last assumption consists in taking the simple average of these lines to get a single number, and the result is a vulnerability line of US\$ 10.1 dollar-a-day in 2005 PPP or US\$ 13.2 dollar-a-day in 2011 PPP. Even though the results presented in the next section are based on the US\$ 10 dollar-a-day 2005 PPP vulnerability line, welfare dynamics were estimated using other vulnerability lines for the 30 percent vulnerability index in other years (i.e. US\$ 9 and US\$ 11.3 dollar-a-day 2005 PPP) showing no significant differences.¹⁵

We would like to offer some further reflections about identifying the vulnerability line. This task depends to a large extent on the specific context of the country, and subjective judgment.¹⁶ Thus it can be useful to combine details from both the contextual background, as well as findings from previous studies to construct the vulnerability line. This process ensures that different economic and societal factors are fully taken into account. Our discussion above did so and suggests that a vulnerability index of 30 percent could be appropriate for Colombia. But we will also examine other robustness checks in future research, for example, by investigating mobility patterns when the vulnerability lines are varied.¹⁷

5. Results: “Monetary welfare” dynamics

This section presents mobility estimates for Colombia during four pairs of years (i.e. 2008-2010, 2010-2012, 2012-2014 and 2014-2016) relying on the synthetic panel methodology proposed in Dang and

¹⁵ Transition matrices as well as profiles are available upon request.

¹⁶ Notably, the construction of the poverty line is also full of arbitrary choices, for example over the composition of the food basket or how to add the non-food expenditure component.

¹⁷ In a quick review, the transition matrices are not statistically different if we use the vulnerability line associated to a vulnerability index of 32 and 28 percent. All these transition matrices are available upon request.

Lanjouw (2013), Dang et al. (2014) and Dang and Lanjouw (2016). In addition, these estimates use cross sectional data from the GEIH during the period 2008–2016 based on the official poverty line and a vulnerability line of US\$ 10 dollar-a-day in 2005 PPP (i.e. US\$ 13.2 dollar-a-day 2011 PPP), and considering the partial correlation coefficient computed with pseudo-panel data for each couple of years from 2008 to 2016. The partial correlation coefficients were 0.51, 0.58, 0.59 and 0.59 for 2008-2010, 2010-2012, 2012-2014 and 2014-2016 respectively.

The estimates in **Table 5** suggest that roughly 56 percent of the population remain in the same income categories (i.e., the sum of the cells in the main diagonal of the matrices), 20 percent experience downward mobility and 24 percent experienced upward mobility. Results show that the estimated poverty rate fell by 14.6¹⁸ percent during 2008-2010, 9.4 percent in 2010-2012, 6.4 percent in 2012-2014, and 3 percent in 2014-2016; while the middle class grew in all these periods except for the last period (2014-2016). While 25.7 percent of the population was vulnerable in 2008, the vulnerable group increased by 14 percent, 3.7 and 2.7 percent between 2008-2010, 2010-2012 and 2014-2016, respectively; it fell 4 percent between 2012-2014.

The increase of the middle class is mainly explained by upward mobility of the population in the vulnerable group. In 2010, around 60 percent of the newcomers to the middle-class came from the vulnerable group. In 2016, 73 percent of the newcomers were vulnerable in 2014. Moreover, the increase of the vulnerable group during the last years is explained by a downward mobility of the middle-class. More precisely, the percentage of poor people joining the vulnerable group decreases each year. In 2010, around 59 percent of the newcomers to the vulnerable group were poor in 2008, but this percentage fell to 48 percent in 2016.

It is also possible to break down the analysis by subgroups based on observable population characteristics such as gender and education level. When comparing welfare dynamics among male and female household heads, we observe that overall mobility is similar across genders and over time (**Figure 4-7**). However, female household heads are slightly less likely to escape poverty in every period than their male counterparts. Thus, poverty shrinks by 15.1, 9.4, 6.3 and 3 percentage points for male household heads; while only by 13.6, 9.2, 6.7 and 3 percentage points for female household heads during 2008-10, 2010-12, 2012-14 and 2014-16, respectively. Despite the fact that poverty rates among females tend to be higher than for males (on average 4 percentage points higher), welfare dynamics are similar across both groups.

¹⁸ This number represents the percentage variation of the poor population between 2008 and 2010. As shown in Table 6, in 2008, 43 percent of the population were poor. By 2010, the poor population was 36.7 percent of the whole population. The percentage variation between the poor population on 2008 and 2010 is $(36.7-43)/43 = -14.6\%$.

Figures 4 to 7 show the welfare dynamics for groups with different levels of education attainment. In terms of welfare dynamics, the population with the highest level of education (tertiary) remained significantly more immobile during the period of analysis (i.e. on average approximately 70 percent of the population stayed in a similar income category between pairs of years from 2008 to 2016). In addition, the population with primary education, middle school and secondary education showed similar levels of immobility (i.e. on average 55 percent of households remained immobile across pairs of years). Moreover, the group of households whose heads were uneducated showed lower overall mobility (i.e. on average almost 60 percent of the population remained in the same income category between pairs of years).

The rate at which the poor could escape poverty and move to vulnerability is shown by **Figure 8**. This presents not only the overall rates but also by different observable characteristics of the head of the household such as gender, level of education and age. For instance, it shows that the upward mobility from poverty¹⁹ towards vulnerability is slightly higher among households headed by males than by females. More importantly, these rates increase with the level of education of the household head. In particular, households whose head has no education are substantially less likely to move up the ladder than any other education group. This same figure shows that the highest rates of escaping poverty happened during the 2010-2012 interval. It is relevant to point out that the results found for households who had fallen into poverty mirror those mentioned above. For instance, the likelihood to fall into poverty is higher for less educated household heads, female and younger.²⁰

Finally, the upward mobility from vulnerability to the middle class²¹ classified by gender and education level of the household head is shown in **Figure 9**. Similarly, households where the head is male are slightly more likely to escape vulnerability towards the middle class, while the rate of upward mobility from vulnerability to middle class also increases with the level of education of the head of the household. In addition, note that it is not immediately clear that there was a period where the upward mobility from vulnerability to join the middle class was always higher than in other periods.

¹⁹ The upward mobility from poverty to vulnerable is the ratio between the population who move out of poverty to vulnerability divided by the sum of the population who transition out of poverty (either to vulnerability or middle class) and who remain in poverty.

²⁰ We do not discuss movements from poor to middle class, given the small sample size. Results available upon request.

²¹ The upward mobility from vulnerability to the middle class is the ratio between the population who move out of vulnerability to middle-class divided by the sum of the population who transition out of vulnerability (either to poverty or middle class) and who remain in vulnerability.

6. Final remarks

This study contributes to filling the gap in the empirical literature currently limiting an evidence-based policy design in Colombia, by identifying a vulnerability line for the country when there are no official available panel data and shedding light on the welfare dynamics during the last decade. The identification of a vulnerability line comes near to the identification of the poverty line, irrespective of the method used. Both methods require multiple assumptions and arbitrary decisions to find the threshold. For instance, in the poverty line process several decisions must be made like the definition of a population of reference, or a caloric threshold and a non-allowance budget share, to mention a few. Similarly, a specific vulnerability index or a probability level to fall into poverty is needed to identify the vulnerability line depending on the method used. These processes are not easy and sensitivity analysis is advisable for each decision.

Currently, an important limitation faced by policy makers in Colombia is the lack of official available panel data, for better informed policy design. This study follows Dang and Lanjouw's (2016) method to overcome this difficulty. It also performs several sensitivity analyses to better inform the crucial decision of choosing a vulnerability line. Thus, US\$ 10 dollar-a-day in 2005 PPP (i.e. US\$ 13.2 dollar-a-day in 2011 PPP) is proposed as the vulnerability threshold. Once this line was identified, this paper implements an analysis of the dynamics of welfare. Briefly, the results show that there is clear evidence of an increase in the size of the middle class during the period and the rate of escaping poverty, and moving from vulnerability into the middle class increases with the level of education.

References

- Atkinson, A. and Brandolini, A. (2011). "On the identification of the 'middle class". *ECINEQ 2011-217*. September 2011.
- Birdsall, N., Graham, C. y Pettinato, S. (2000) Stuck in the tunnel: Is globalization muddling the middle class? *Center on Social and Economic Dynamics*, WP 14
- Blackburn, M. y Bloom, D. (1987) Earnings and income inequality in the United States, *Population Development Review*, pp. 575-609
- Bourguignon, F. and Moreno, H. (2015). "On the construction of synthetic panels". *MIMEO. NEUDC*. October 2015.
- Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R. and Viollaz, M. (2015). "Intra-Generational Mobility and Repeated Cross-Sections: A Three-Country Validation Exercise." *Journal of Economic Inequality*, 13 (2): 161–79.
- D'Ambrosio, C., Muliere, P. y Secchi, P. (2002), Income thresholds and income classes, WP, Università Bocconi, Milano, Italy
- Dang, Hai-Anh and Peter Lanjouw. (2013). "Measuring poverty dynamics with synthetic panels based on cross-sections." World Bank Policy Research Working Paper 6540
- . (2016). "Measuring Poverty Dynamics with Synthetic Panels Based on Repeated Cross-Sections." Available at: <http://lacer.lacea.org/handle/123456789/61406>
- . (2017). "Welfare Dynamics Measurement: Two Definitions of a Vulnerability Line and Their Empirical Application". *Review of Income and Wealth*, 63: 633-660.
- Dang, Hai-Anh, Peter Lanjouw, Jill Luoto, and David McKenzie. (2014). "Using Repeated Cross-Sections to Explore Movements in and out of Poverty". *Journal of Development Economics*, 107: 112-128.
- Deaton, A. (1985). "Panel Data from Time Series of Cross-Sections". *Journal of Econometrics*, 30 (1985) 109-126. North-Holland.
- Deaton, A. and Paxson, C. (2004). "Intertemporal Choice and Inequality". *The Journal of Political Economy*, Vol. 102, No. 3 (Jun., 1994), pp. 437-467.
- Fields, G., & Viollaz, M. (2013). "Can the Limitations of Panel Datasets be Overcome by Using Pseudo-Panels to Estimate Income Mobility?" *Universidad Cornell-CEDLAS*.
- Hertova, D., López-Calva, L. F., & Ortiz-Juárez, E. (2010). Bigger... but Stronger? The Middle Class in Chile and Mexico in the Last Decade. *Research for Public Policy, Inclusive Development, ID-02-2010, RBLAC-UNDP, New York*.
- Horrigan, M. y Haugen, S. (1988), The declining middle-class thesis: a sensitivity analysis, *Monthly Labor Review*, 111, 3-13
- Jarque, C. M. and Bera, A. K. (1980). "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". *Economics Letters*. 6 (3): 255–259
- Kosters, M. y Ross, M. (1988) A shrinking middle class? *Public Interest*, 90, pp 3-27
- Lawrence, R. (1984) Sectoral shifts and size of the middle class, *Brookings Review*, pp. 3-11
- Lopez-Calva, L. F., and Ortiz-Juarez, E. (2014). "A Vulnerability Approach to the Definition of the Middle Class." *The Journal of Economic Inequality*, 12(1): 23-47.
- Lucchetti, L. R. (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 8820.
- Pavon, L. and Perez, C. A. (2016). "Medición y Caracterización de la Clase Media en Colombia". *Subdirección de Promoción Social y Calidad de Vida del Departamento Nacional de Planeación*. Mimeo.
- Pencavel, J. (2006). "A Life Cycle Perspective on Changes in Earnings Inequality among Married Men and Women". *The Review of Economics and Statistics*, May 2006, 88(2): 232-242.
- Veeberk, M. (2007). "Pseudo Panels and repeated cross-sections". Chapter prepared for: L. Mátyás and P. Sevestre, eds., (2008), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Springer.

Table 1. Poverty Rates (2008-2016)

| Year | Full Sample | Restricted Sample | Difference |
|-------------|--------------------|--------------------------|-------------------|
| 2008 | 42.2 | 43.3 | -1.16 *** |
| 2009 | 40.4 | 41.4 | -1.01 *** |
| 2010 | 37.3 | 38.0 | -0.72 *** |
| 2011 | 34.2 | 34.9 | -0.70 *** |
| 2012 | 32.9 | 33.2 | -0.30 *** |
| 2013 | 30.7 | 31.0 | -0.33 *** |
| 2014 | 28.6 | 28.7 | -0.04 |
| 2015 | 27.9 | 27.8 | 0.16 |
| 2016 | 28.1 | 27.6 | 0.45 *** |

Note: The table shows estimates of the moderate poverty rate in Colombia using the *full sample* of the GEIH and a *restricted sample* based on head of households aged 25 to 60 years in the first survey (GEIH-2008) and adjusted accordingly for later waves of the survey.

Source: Own estimations based on GEIH from 2008 to 2016.

Table 2. Demographic Composition of the Sample (2008-2016)

| Characteristics of Head | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|--------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Cohort group | | | | | | | | | |
| 1974 and earlier | 22.2 (0.4) | 23.7 (0.4) | 25.5 (0.4) | 26.7 (0.4) | 28.2 (0.5) | 29.2 (0.5) | 30.4 (0.5) | 31.2 (0.5) | 31.7 (0.5) |
| 1964-1973 | 34.1 (0.5) | 34.4 (0.5) | 34.1 (0.5) | 34.0 (0.5) | 33.4 (0.5) | 33.1 (0.5) | 33.4 (0.5) | 33.2 (0.5) | 33.6 (0.5) |
| 1954-1963 | 31.2 (0.5) | 29.7 (0.5) | 28.7 (0.5) | 27.9 (0.4) | 27.4 (0.4) | 26.7 (0.4) | 26.1 (0.4) | 25.4 (0.4) | 24.7 (0.4) |
| 1948-1953 | 12.5 (0.3) | 12.2 (0.3) | 11.7 (0.3) | 11.5 (0.3) | 11.0 (0.3) | 11.0 (0.3) | 10.1 (0.3) | 10.2 (0.3) | 10.0 (0.3) |
| Gender | | | | | | | | | |
| Female | 24.8 (0.4) | 26.1 (0.4) | 27.1 (0.4) | 28.3 (0.5) | 29.7 (0.5) | 30.5 (0.5) | 32.0 (0.5) | 32.7 (0.5) | 32.9 (0.5) |
| Male | 75.2 (0.4) | 73.9 (0.4) | 72.9 (0.4) | 71.7 (0.5) | 70.3 (0.5) | 69.5 (0.5) | 68.0 (0.5) | 67.3 (0.5) | 67.1 (0.5) |
| Education level | | | | | | | | | |
| No education | 6.6 (0.2) | 6.9 (0.3) | 6.9 (0.3) | 6.9 (0.3) | 7.1 (0.3) | 6.9 (0.3) | 6.7 (0.2) | 6.6 (0.2) | 6.6 (0.2) |
| Primary education | 39.8 (0.5) | 40.3 (0.5) | 39.6 (0.5) | 39.2 (0.5) | 38.7 (0.5) | 37.9 (0.5) | 37.5 (0.5) | 37.6 (0.5) | 36.9 (0.5) |
| Middle school | 16.6 (0.4) | 16.7 (0.4) | 16.3 (0.4) | 16.0 (0.4) | 15.9 (0.4) | 15.6 (0.4) | 15.4 (0.4) | 15.1 (0.4) | 14.9 (0.4) |
| Secondary education | 20.8 (0.4) | 20.7 (0.4) | 20.8 (0.4) | 21.4 (0.4) | 21.5 (0.4) | 21.3 (0.4) | 21.4 (0.4) | 22.2 (0.4) | 22.7 (0.4) |
| Higher education | 16.2 (0.4) | 15.4 (0.4) | 16.4 (0.4) | 16.4 (0.4) | 16.8 (0.4) | 18.3 (0.4) | 19.0 (0.4) | 18.6 (0.4) | 18.8 (0.4) |
| Household size | | | | | | | | | |
| 1-2 people | 9.4 (0.3) | 9.9 (0.3) | 10.4 (0.3) | 10.7 (0.3) | 11.3 (0.3) | 11.8 (0.3) | 12.8 (0.3) | 13.2 (0.3) | 13.8 (0.3) |
| 3-4 people | 42.4 (0.5) | 43.3 (0.5) | 43.5 (0.5) | 43.6 (0.5) | 43.6 (0.5) | 44.5 (0.5) | 44.3 (0.5) | 44.1 (0.5) | 44.0 (0.5) |
| 5-6 people | 32.5 (0.5) | 31.9 (0.5) | 31.6 (0.5) | 31.6 (0.5) | 30.5 (0.5) | 29.7 (0.5) | 29.6 (0.5) | 29.7 (0.5) | 28.9 (0.5) |
| more than 6 people | 15.8 (0.4) | 15.0 (0.4) | 14.5 (0.4) | 14.2 (0.3) | 14.6 (0.4) | 14.0 (0.3) | 13.2 (0.3) | 13.0 (0.3) | 13.4 (0.3) |

Note: Estimates based on sample of head of households aged 25 to 60 years in the first survey (GEIH-2008) and adjusted accordingly for later waves. The estimates represent the share of the household heads. Standard deviations in parenthesis.

Table 3. Vulnerability Lines and given Vulnerability Indexes for Colombia (2008-2010)

| No | Vulnerability index (%) ^a | Vulnerability line (\$ real COP, Bogota 2008) ^b | Vulnerability line (\$PPP 2011) ^c | Increase with comparison to poverty line (%) ^d | Pop. share with consumption above poverty line but less than V-line (%) ^e |
|-----|--------------------------------------|--|--|---|--|
| (1) | (2) | (3) | (4) | (5) | (6) |
| 1 | 35 | 238603 | 7.11 | 22 | 7 |
| 2 | 34 | 230803 | 6.88 | 18 | 6 |
| 3 | 33 | 226903 | 6.76 | 16 | 5 |
| 4 | 32 | 371203 | 11.06 | 90 | 19 |
| 5 | 31 | 425803 | 12.68 | 118 | 23 |
| 6 | 30 | 488203 | 14.54 | 149 | 27 |
| 7 | 29 | 558403 | 16.63 | 185 | 30 |
| 8 | 28 | 648103 | 19.31 | 231 | 34 |
| 9 | 27 | 757303 | 22.56 | 287 | 37 |
| 10 | 26 | 897703 | 26.74 | 359 | 40 |
| 11 | 25 | 1084903 | 32.32 | 454 | 44 |
| 12 | 24 | 1338403 | 39.87 | 584 | 47 |
| 13 | 23 | 1769603 | 52.72 | 804 | 50 |
| 14 | 22 | 2583003 | 76.95 | 1220 | 53 |

^a The vulnerability index is the share of the vulnerable population in the first period that becomes poor in the second period. The vulnerable population is understood here as those people with income per-capita above the poverty line, but below the corresponding vulnerability line.

^b The vulnerability line corresponds to the monetary-value per person, below such value and above the poverty line, for which the probability of falling into poverty in the second period is the associated vulnerability index.

^d Since vulnerability line is above the poverty line, this column reflects the relative increases of the vulnerability line from the poverty line (of \$195703 at real COP, Bogota 2008). The incremental value for iteration is 2 percent of the poverty line (that is \$3915 real COP, Bogota 2008).

^e The population share in this case refers to the percentage of people with consumption below the vulnerability line and above the poverty line in both periods.

Source: Own estimations based on GEIH from 2008 to 2016.

Table 4. Vulnerability Lines using Pseudo-Lopez Calva Method (2008-2016)

| Year | Vulnerability line (\$ real COP, Bogota 2008) ^a | Vulnerability line (\$PPP 2011) ^b | Increase with comparison to poverty line (%) ^c |
|----------------|--|---|---|
| <i>Model 1</i> | | | |
| 2008-2010 | 519671 | 15.48 | 166 |
| 2010-2012 | 513082 | 15.28 | 162 |
| 2012-2014 | 524594 | 15.63 | 168 |
| 2014-2016 | 597737 | 17.81 | 205 |
| <i>Model 2</i> | | | |
| 2008-2010 | 401966 | 11.97 | 105 |
| 2010-2012 | 412978 | 12.30 | 111 |
| 2012-2014 | 386059 | 11.50 | 97 |
| 2014-2016 | 457724 | 13.64 | 134 |
| <i>Model 3</i> | | | |
| 2008-2010 | 384313 | 11.45 | 96 |
| 2010-2012 | 393411 | 11.72 | 101 |
| 2012-2014 | 404297 | 12.04 | 107 |
| 2014-2016 | 446504 | 13.30 | 128 |
| <i>Model 4</i> | | | |
| 2008-2010 | 385275 | 11.48 | 97 |
| 2010-2012 | 398487 | 11.87 | 104 |
| 2012-2014 | 409832 | 12.21 | 109 |
| 2014-2016 | 442236 | 13.17 | 126 |

^a The vulnerability line corresponds to the monetary-value per person, below such value and above the poverty line, for which the probability of falling into poverty in the second period is 10%.

^c Since vulnerability line is above the poverty line, this column reflects the relative increases of the vulnerability line from the poverty line (of \$195703 at real COP, Bogota 2008).

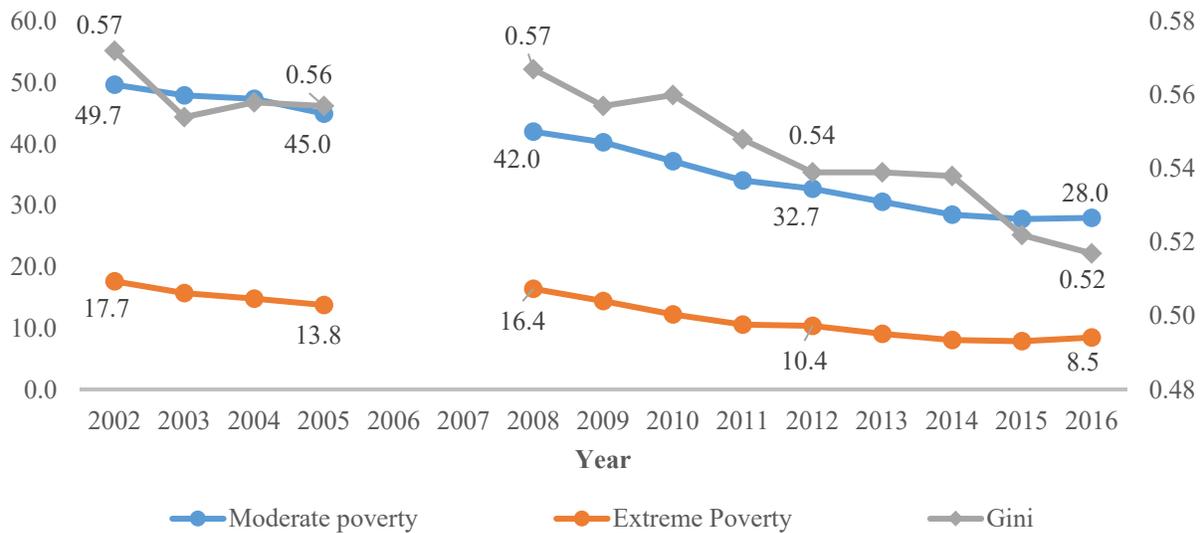
Source: Own estimations based on GEIH from 2008 to 2016.

Table 5. Welfare Dynamics
(Periods 2008-10, 2010-12, 2012-14, 2014-16)

| | | 2010 | | | |
|------|--------------|-----------------|-----------------|-----------------|-------|
| | | Poverty | Vulnerable | Middle class | Total |
| 2008 | Poverty | 26.7 (0.100) | 11.8 (0.021) | 4.6 (0.004) | 43.0 |
| | Vulnerable | 7.3 (0.016) | 9.8 (0.009) | 8.6 (0.019) | 25.7 |
| | Middle class | 2.7 (0.004) | 7.7 (0.011) | 20.8 (0.113) | 31.3 |
| | Total | 36.7 | 29.3 | 34.0 | |
| | | 2012 | | | |
| | | Poverty | Vulnerable | Middle class | Total |
| 2010 | Poverty | 23.1 (0.099) | 10.5 (0.027) | 3.7 (0.005) | 37.3 |
| | Vulnerable | 8.2 (0.020) | 11.6 (0.015) | 9.5 (0.015) | 29.3 |
| | Middle class | 2.6 (0.003) | 8.2 (0.014) | 22.6 (0.134) | 33.4 |
| | Total | 33.8 | 30.4 | 35.8 | |
| | | 2014 | | | |
| | | Poverty | Vulnerable | Middle class | Total |
| 2012 | Poverty | 20.8 (0.084) | 9.5 (0.024) | 3.7 (0.005) | 34.1 |
| | Vulnerable | 8.4 (0.018) | 11.5 (0.014) | 10.4 (0.011) | 30.3 |
| | Middle class | 2.7 (0.003) | 8.1 (0.012) | 24.8 (0.122) | 35.6 |
| | Total | 31.9 | 29.1 | 38.9 | |
| | | 2016 | | | |
| | | Poverty | Vulnerable | Middle class | Total |
| 2014 | Poverty | 19.2 (0.073) | 9.0 (0.023) | 3.4 (0.006) | 31.6 |
| | Vulnerable | 8.3 (0.018) | 11.2 (0.016) | 9.5 (0.008) | 29.0 |
| | Middle class | 3.2 (0.003) | 9.6 (0.012) | 26.6 (0.118) | 39.4 |
| | Total | 30.7 | 29.8 | 39.5 | |

Source: Own estimations based on GEIH from 2008 to 2016.

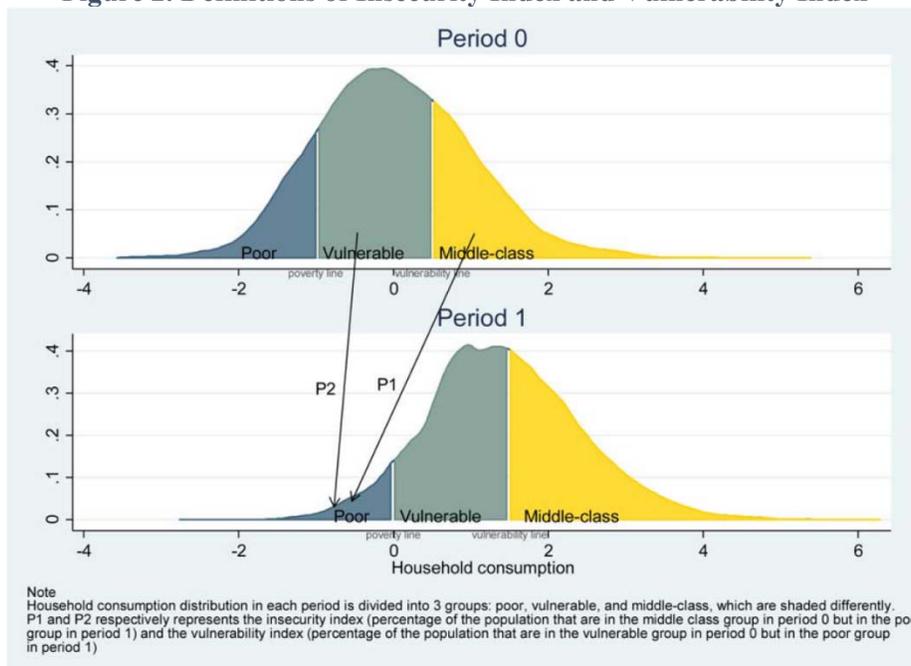
Figure 1. Evolution of Poverty Rates in Colombia (2002-2016)



Note: Monetary poverty estimates are based on the official poverty line. The *Misión de Empalme de las Cifras de Pobreza y Mercado Laboral* (MESEP) committee decided not to report monetary poverty estimates for 2006 and 2007 given the methodological changes that took place in those years. The committee deemed that only estimates based on the 2002–05 and 2008–15 series are comparable.

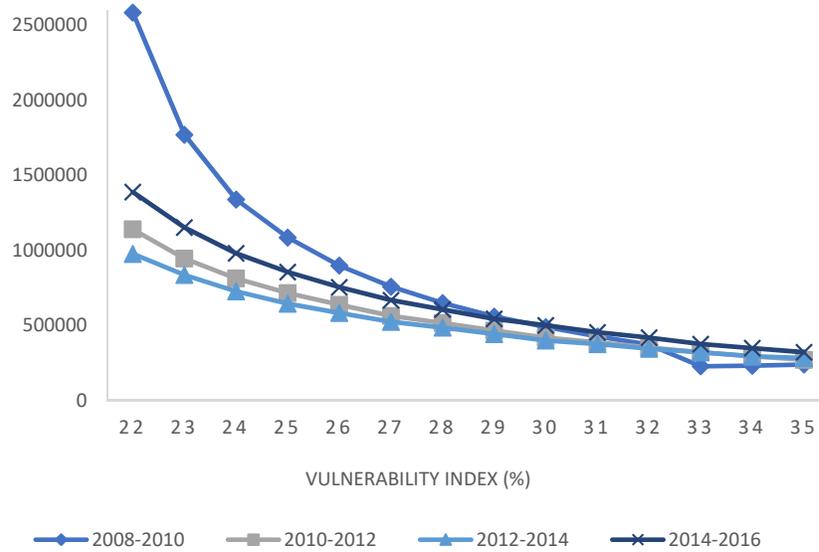
Source: Own elaboration based on data of the *Departamento Administrativo Nacional de Estadística* (National Administrative Department of Statistics of Colombia, DANE).

Figure 2. Definitions of Insecurity Index and Vulnerability Index



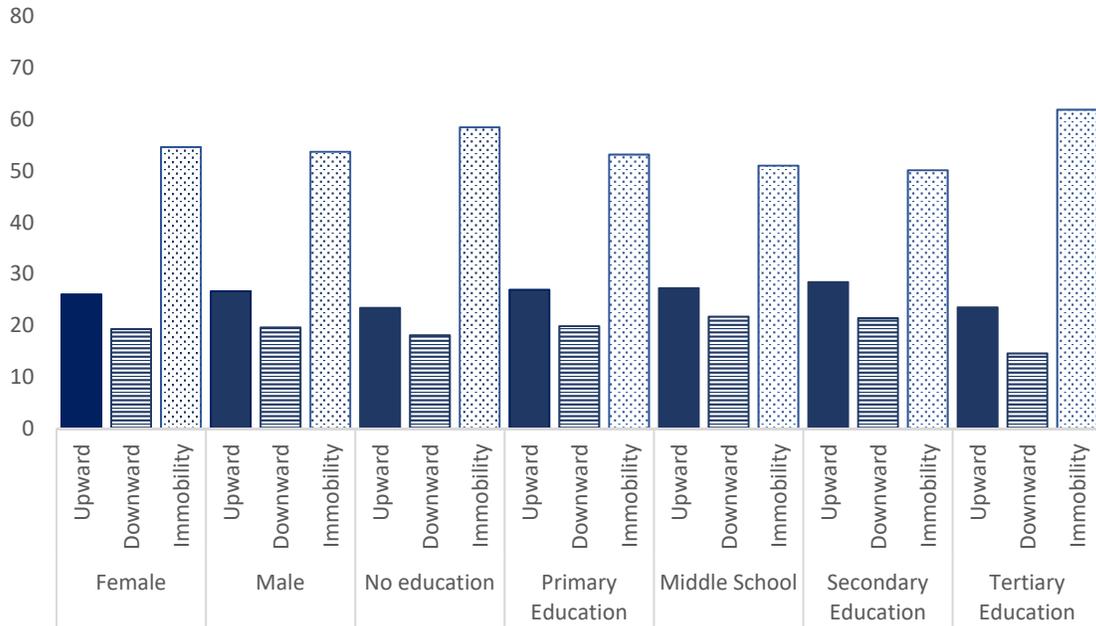
Source: Dang and Lanjouw (2017).

Figure 3. Vulnerability lines in Colombian pesos for different base years (Periods 2008-10, 2010-12, 2012-14 and 2014-16)



Source: Own estimations based on GEIH from 2008 to 2016.

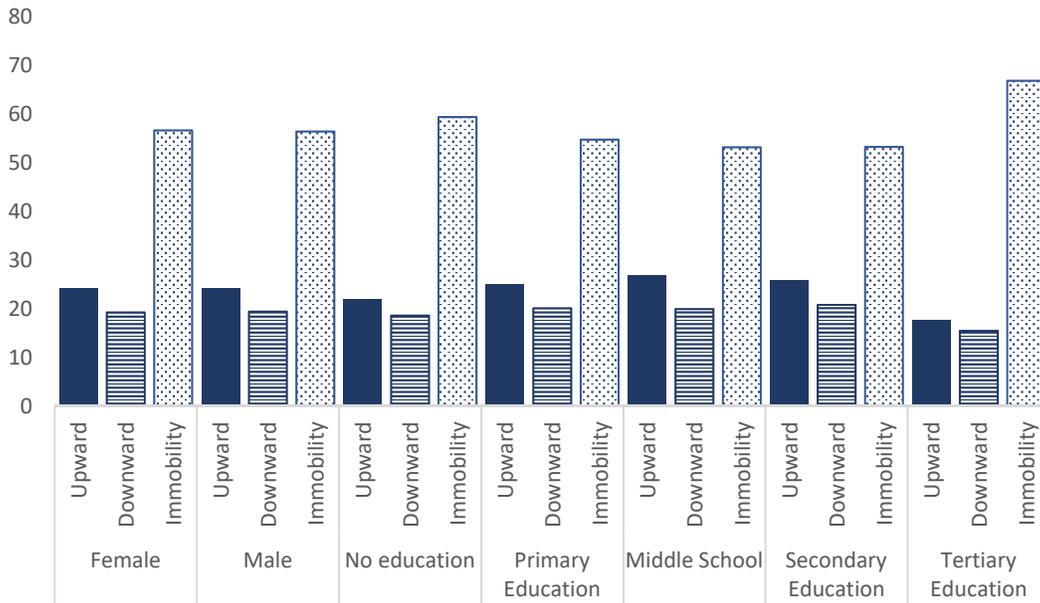
Figure 4: Welfare mobility by gender, and Education levels for period 2008-10



Note: upward movers: proportion of poor or vulnerable population in 1st period that move up one or two income categories in 2nd period. Downward movers: proportion of vulnerable or middle-class population in 1st period that move down one or two income categories in 2nd period. Vulnerability line is set at US\$13.2 dollar –a-day 2011 PPP.

Source: Own estimations based on GEIH from 2008 to 2010.

Figure 5: Welfare mobility by gender, and Education levels for period 2010-12

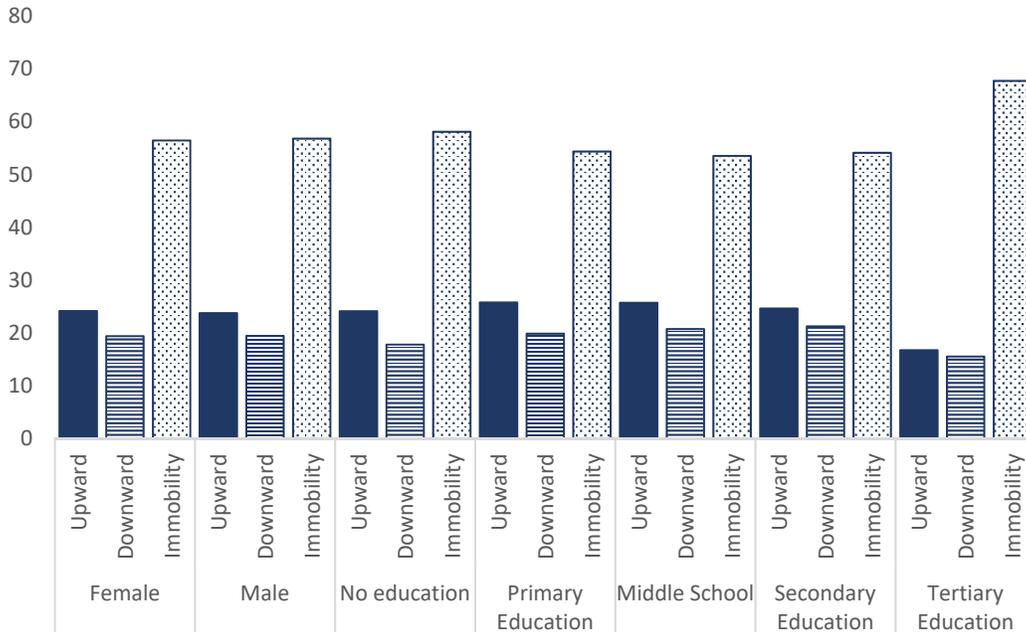


Note: Upward movers: proportion of poor or vulnerable population in 1st period that move up one or two income categories in 2nd period. Downward movers: proportion of vulnerable or middle-class population in 1st period that move down one or two income categories in 2nd period.

Vulnerability line is set at US\$13.2 dollar –a-day 2011 PPP.

Source: Own estimations based on GEIH from 2010 to 2012.

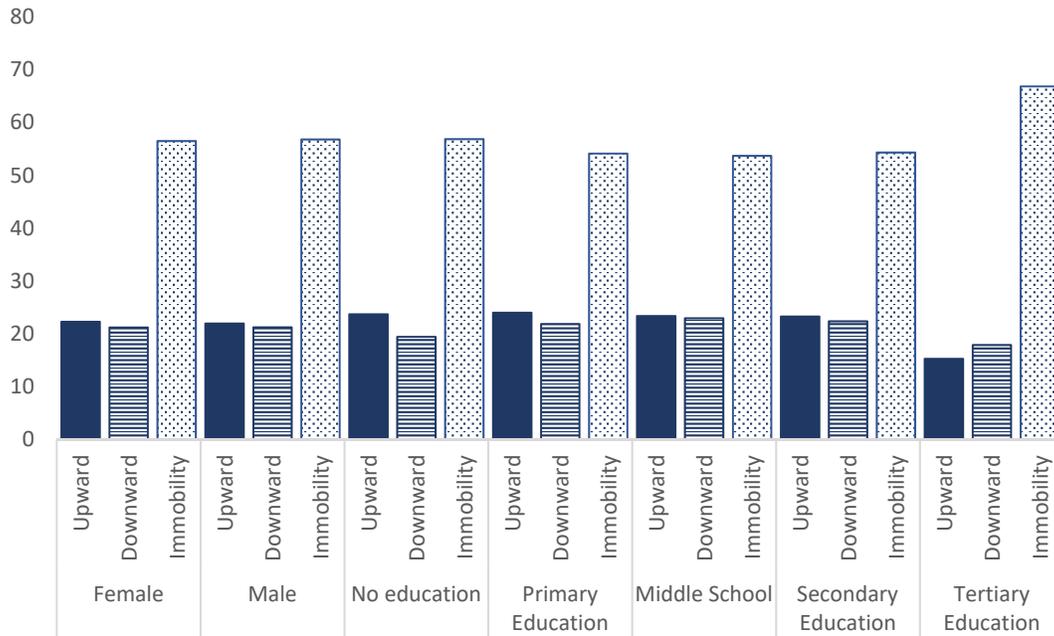
Figure 6: Welfare mobility by gender, and Education levels for period 2012-14



Note: Upward movers: proportion of poor or vulnerable population in 1st period that move up one or two income categories in 2nd period. Downward movers: proportion of vulnerable or middle-class population in 1st period that move down one or two income categories in 2nd period.

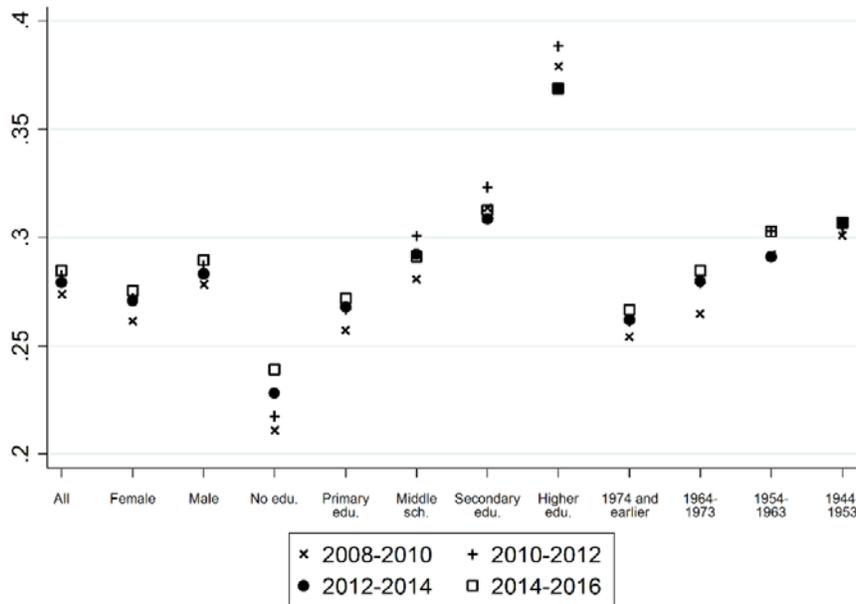
Vulnerability line is set at US\$13.2 dollar –a-day 2011 PPP.

Figure 7: Welfare mobility by gender, and Education levels for period 2014-16



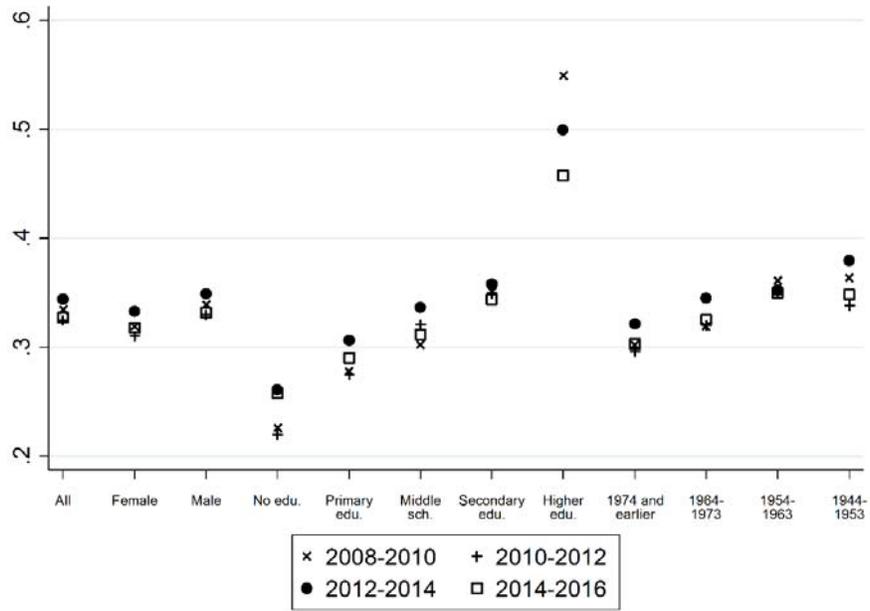
Note: Upward movers: proportion of poor or vulnerable population in 1st period that move up one or two income categories in 2nd period. Downward movers: proportion of vulnerable or middle-class population in 1st period that move down one or two income categories in 2nd period. Vulnerability line is set at US\$13.2 dollar –a-day 2011 PPP.

Figure 8. Upward Mobility from Poverty to Vulnerability (Periods 2008-10, 2010-12, 2012-14 and 2014-16)



Source: Own estimations based on GEIH from 2008 to 2016.

**Figure 9. Upward Mobility from Vulnerability to Middle Class
(Periods 2008-10, 2010-12, 2012-14 and 2014-16)**



Source: Own estimations based on GEIH from 2008 to 2016.

Appendix A: Welfare dynamics with a poverty line

When there is only one poverty line z of interest (using both incomes $y_{i,j}$ and the poverty line z are expressed in real terms), it is possible to represent the four relevant states using a 2 by 2 matrix, such as the one depicted in

Table A. 1 below:

(a) $\Pr(y_1 < z \text{ and } y_2 < z)$, which represents the percentage of households that were poor in the first and the second round,

(b) the percentage of households that are poor in the first round and non-poor in the second round $\Pr(y_1 < z \text{ and } y_2 > z)$.

(c) the percentage of households that are non-poor in the first round and poor in the second round : $\Pr(y_2 < z \text{ and } y_1 > z)$.

(d) the percentage of households that are non-poor in both survey rounds, $\Pr(y_2 > z \text{ and } y_1 > z)$.

However, notice that when only repeated cross-sections are available, it is not straightforward to construct the transitions in

Table A. 1, since it is not possible to observe the same household in multiple periods, as would occur with panel data.

Then, for example, the percentage of households that are poor in the first round and non-poor in the second round could be estimated using the following probability: $\Pr(y_1 < z \text{ and } y_2 > z)$. The prime difficulty with repeated cross-sections is that the researcher is not able to observe the values of y_{i1} and y_{i2} for the same household in multiple periods. However, it is possible to write the previous probability as a function of the joint distribution of the error terms ε_{i1} and ε_{i2} , capturing the correlation of those parts of the household consumption in the two periods which are not explained by the household characteristics $x_{i,1}$ and $x_{i,2}$:

$$\Pr(\varepsilon_{i1} < z - \beta_1'x_{i,1} \text{ and } \varepsilon_{i2} > z - \beta_2'x_{i,2})$$

Importantly, it is possible to operationalize the previous expression relying on a bivariate normal distribution and using $\Phi_2(\cdot)$ to represent the bivariate normal cumulative distribution function (cdf) as:

$$P(y_{i,1} < z \text{ and } y_{i,2} > z) = \Phi_2\left(\frac{z - \beta_1'x_{i,1}}{\sigma_{\varepsilon_1}}, \frac{z - \beta_2'x_{i,2}}{\sigma_{\varepsilon_2}}, -\rho\right)$$

A key element in the analysis of income mobility is the estimation of the correlation coefficient ρ , that is likely to be non-negative, using one of the following alternatives:

i. First, the simplest case occurs if ρ is known, since the estimation of the bivariate normal distribution becomes relatively straightforward. However, this is not the typical case, since the real value

of ρ is usually unknown in many contexts, such as in the case of Colombia (when we do not have panel data).

ii. Second, we can obtain the upper and lower bounds of mobility by assuming minimum and maximum values for the correlation, for instance starting with $\rho = 0$ and $\rho = 1$. In the first case ($\rho = 0$), the researcher is implicitly assuming that there is zero correlation between the error terms, thus income prediction for the first round is done by randomly drawing with replacement from the empirical distribution of the first-round estimated residuals for each household i in the second round. In the second case, when $\rho = 1$, the implicit assumption is that the correlations of the idiosyncratic shocks are perfect and positive, adding more “persistence” and “stickiness” to the vector of income.

iii. Third, we can identify a range of values for ρ , from a group of similar comparable countries with actual panel data.²² This method would also allow to refine the bounds for the higher and lower values that ρ could take. Moreover, instead of a range of values we could adopt a single value for ρ based on information from comparable sources.

iv. Finally, it is possible to obtain an approximation of this ρ based on asymptotic theory following the approach proposed by Dang and Lanjouw (2016). The procedure implies aggregating all the variables at a cohort level, and estimating the following cohort-level equation:

$$y_{cj} = \beta_j' x_{cj} + \varepsilon_{cj}$$

Then the partial correlation coefficient can be estimated as follows,

$$\rho_{\varepsilon_{c1}\varepsilon_{c2}} = \frac{cov(\varepsilon_{c1}, \varepsilon_{c2})}{\sqrt{var(\varepsilon_{c1})var(\varepsilon_{c2})}}$$

Notice that the estimates of β_j correspond to the linear projection of household income—or consumption—on household characteristics aggregated at a cohort level for survey round $j=1,2$.

²² Lucchetti (2017) provides evidence from different countries: “Residual correlation from panel data estimates are on average about 0.65 in Dang and Lanjouw (2013), about 0.50 in Dang et al. (2014), and about 0.45 in Cruces et al. (2015)... The correlation is about 0.50 in Peru between 2007 and 2011.”

Appendix B: Welfare dynamics with a poverty and a vulnerability line

The synthetic panel methodology can be used to analyze the mobility between more than two states i.e., when there is more than one relevant line. For instance, assuming that the relevant welfare aggregate is based on monetary income and that there are two lines, a poverty line (like the one described in the previous appendix) and a vulnerability line, such that the combination of these two lines would create three possible states for each household: (i) poor, if the household income lies below the poverty line; (ii) vulnerable, when the income is above the poverty line but below the vulnerability line; and (iii) middle class, when the average income of the household is higher than the vulnerability line.

Let $x_{i,j}$ be a vector of household characteristics observed in survey round $j = \{1,2\}$ for household of individual $i = \{1,2, \dots, N\}$, with N representing the sample size. In addition, continue to let y_{ij} represent the income for household i in round $j = \{1,2\}$.

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

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

Notice that this framework requires two assumptions:

(1) The underlying populations being sampled in both surveys are identical such that their **time-invariant** characteristics remain the same over time. This implies that the conditional distribution of expenditure or income in each period is identical whether it is conditional on given household characteristics in the first or second period. In absence of panel data this ensures that we can use time-invariant household characteristics that are observed in both survey rounds to obtain predicted household incomes.

(2) Both $\varepsilon_{i,1}$ and $\varepsilon_{i,2}$ have a bivariate normal distribution with non-negative correlation coefficient ρ and standard deviations σ_{ε_1} and σ_{ε_2} , respectively.²³

The first assumption can be tested empirically using a mean-comparison test to assess if the observable characteristics assumed to be time invariable are statistically different across years. In addition, the second assumption on the normality of the residuals can also be tested using standard statistical procedures.

Notice that in contrast with the analysis of transitions in and out of poverty, there are now two relevant lines, a poverty line in period $j = \{1,2\}$ represented by z_j and a vulnerability line depicted by v_j . For example, the percentage of households that are poor in the first period and vulnerable in the second period is given by:

²³ Note that the assumption on the normality of the residuals can be analyzed graphically by plotting the kernel distribution of (log) income residuals for each round and comparing them with the normal distribution or analytically using Skewness and Kurtosis tests, as well as Shapiro-Wilk normality tests (Bourguignon and Moreno, 2015) or the Jarque and Bera (1980) test.

$$P(y_{i1} < z_1 \quad \text{and} \quad z_2 < y_{i2} < v_2)$$

Given the two relevant lines and three alternative states of interest, is possible to represent mobility as a three by three matrix with nine potential scenarios of income mobility, such as the example in Table B.1. One of the useful properties of the matrix in Table B.1 is that it allows to directly establish income immobility by summing up the cells on the main diagonal (which correspond to the share of households who remain in the same state in the initial and final periods). Similar to the discussion in section 2.1, it is possible to express the transition probabilities as functions of the joint distribution of the error terms:

(a) Poor-Poor (P_1):

$$P_1 = P(y_{i1} < z_1 \quad \text{and} \quad y_{i2} < z_2) = \Phi_2\left(\frac{z_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right)$$

(b) Poor-Vulnerable (P_2):

$$\begin{aligned} P_2 &= P(y_{i1} < z_1 \quad \text{and} \quad z_2 < y_{i2} < v_2) \\ &= \Phi_2\left(\frac{z_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right) - \Phi_2\left(\frac{z_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right) \end{aligned}$$

(c) Poor-Middle Class (P_3):

$$P_3 = P(y_{i1} < z_1 \quad \text{and} \quad y_{i2} > v_2) = \Phi_2\left(\frac{z_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, -\frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, -\rho\right)$$

(d) Vulnerable-Poor (P_4):

$$P_4 = P(z_1 < y_{i1} < v_1 \quad \text{and} \quad y_{i2} < z_2) = \Phi_2\left(\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right) - \Phi_2\left(\frac{z_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right)$$

(e) Vulnerable-Vulnerable (P_5):

$$\begin{aligned} P_5 &= P(z_1 < y_{i1} < v_1 \quad \text{and} \quad z_2 < y_{i2} < v_2) \\ &= P(y_{i1} < v_1 \quad \text{and} \quad z_2 < y_{i2} < v_2) - P(y_{i1} < z_1 \quad \text{and} \quad z_2 < y_{i2} < v_2) \\ &= \Phi_2\left(\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right) - \Phi_2\left(\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right) - P_2 \end{aligned}$$

(f) Vulnerable-Middle Class (P_6):

$$\begin{aligned} P_6 &= P(z_1 < y_{i1} < v_1 \quad \text{and} \quad y_{i2} > v_2) \\ &= \Phi_2\left(\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, -\frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, -\rho\right) - \Phi_2\left(\frac{z_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, -\frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, -\rho\right) \end{aligned}$$

(g) Middle Class-Poor (P_7):

$$P_7 = P(v_1 < y_{i1} \quad \text{and} \quad y_{i2} < z_2) = \Phi_2\left(-\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, -\rho\right)$$

(h) Middle Class-Vulnerable (P_8):

$$\begin{aligned} P_8 &= P(y_{i1} > v_1 \quad \text{and} \quad z_2 < y_{i2} < v_2) \\ &= \Phi_2\left(-\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, \frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, -\rho\right) - \Phi_2\left(-\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, -\frac{z_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, -\rho\right) \end{aligned}$$

(i) Middle Class-Middle Class (P_9):

$$P_7 = P(y_{i1} > v_1 \text{ and } y_{i2} > v_2) = \Phi_2\left(-\frac{v_1 - \beta_1'x_{i2}}{\sigma_{\varepsilon_1}}, -\frac{v_2 - \beta_2'x_{i2}}{\sigma_{\varepsilon_2}}, \rho\right)$$

Table A. 1. Basic 2x2 Transition Matrix

| | | Final Period | |
|----------------|---------------------------|-----------------------|---------------------------|
| | | Poor ($y_2 < z$) | Non-Poor ($y_2 > z$) |
| Initial Period | Poor ($y_1 < z$) | (a) | (b) |
| | Non-Poor ($y_1 > z$) | (c) | (d) |

Source: Own elaboration

Table B.1. The 3x3 Transition Matrix

| | | Final Period | | |
|----------------|--------------|--------------|------------|--------------|
| | | Poor | Vulnerable | Middle class |
| Initial Period | Poor | (a) | (b) | (c) |
| | Vulnerable | (d) | (e) | (f) |
| | Middle class | (g) | (h) | (i) |

Source: Own elaboration

Appendix C: Additional Tables

Table C.1. Differences in Gender of Household Head (2008-2016)

| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|-----------------------|
| 2008 | 0.0000 (.) | 0.0124*** (0.0031) | 0.0232*** (0.0031) | 0.0353*** (0.0032) | 0.0490*** (0.0033) | 0.0572*** (0.0032) | 0.0719*** (0.0033) | 0.0785*** (0.0033) | 0.0804*** (0.0033) |
| 2009 | -0.0124*** (0.0031) | 0.0000 (.) | 0.0108*** (0.0031) | 0.0229*** (0.0032) | 0.0366*** (0.0033) | 0.0448*** (0.0032) | 0.0595*** (0.0032) | 0.0661*** (0.0033) | 0.0680*** (0.0033) |
| 2010 | -0.0232*** (0.0031) | -0.0108*** (0.0031) | 0.0000 (.) | 0.0121*** (0.0032) | 0.0258*** (0.0033) | 0.0340*** (0.0032) | 0.0487*** (0.0032) | 0.0553*** (0.0033) | 0.0572*** (0.0033) |
| 2011 | -0.0353*** (0.0032) | -0.0229*** (0.0032) | -0.0121*** (0.0032) | 0.0000 (.) | 0.0137*** (0.0033) | 0.0219*** (0.0033) | 0.0366*** (0.0033) | 0.0432*** (0.0033) | 0.0451*** (0.0033) |
| 2012 | -0.0490*** (0.0033) | -0.0366*** (0.0033) | -0.0258*** (0.0033) | -0.0137*** (0.0033) | 0.0000 (.) | 0.0082** (0.0034) | 0.0229*** (0.0034) | 0.0295*** (0.0035) | 0.0314*** (0.0034) |
| 2013 | -0.0572*** (0.0032) | -0.0448*** (0.0032) | -0.0340*** (0.0032) | -0.0219*** (0.0033) | -0.0082** (0.0034) | 0.0000 (.) | 0.0147*** (0.0033) | 0.0213*** (0.0034) | 0.0232*** (0.0034) |
| 2014 | -0.0719*** (0.0033) | -0.0595*** (0.0032) | -0.0487*** (0.0032) | -0.0366*** (0.0033) | -0.0229*** (0.0034) | -0.0147*** (0.0033) | 0.0000 (.) | 0.0066* (0.0034) | 0.0085** (0.0034) |
| 2015 | -0.0785*** (0.0033) | -0.0661*** (0.0033) | -0.0553*** (0.0033) | -0.0432*** (0.0033) | -0.0295*** (0.0035) | -0.0213*** (0.0034) | -0.0066* (0.0034) | 0.0000 (.) | 0.0019 (0.0035) |
| 2016 | -0.0804*** (0.0033) | -0.0680*** (0.0033) | -0.0572*** (0.0033) | -0.0451*** (0.0033) | -0.0314*** (0.0034) | -0.0232*** (0.0034) | -0.0085** (0.0034) | -0.0019 (0.0035) | 0.0000 (.) |

Standard errors in parenthesis, * p<0.1 ** p<0.05 *** p<0.01

Note: Each box presents the differences in the percentage of female household heads between two different given years. As the grey becomes darker, the difference on the percentage of female household heads between the two given years becomes more statistically significant.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.2. Differences in the Education Level of the Household Head - No Education (2008-2016)

| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|------|----------------------|---------------------|---------------------|---------------------|-----------------------|---------------------|---------------------|----------------------|----------------------|
| 2008 | 0.0000 (.) | -0.0033 (0.0024) | -0.0034 (0.0023) | -0.0036 (0.0024) | -0.0053** (0.0025) | -0.0031 (0.0024) | -0.0011 (0.0024) | -0.0002 (0.0024) | 0.0001 (0.0024) |
| 2009 | 0.0033 (0.0024) | 0.0000 (.) | -0.0001 (0.0023) | -0.0003 (0.0024) | -0.0020 (0.0024) | 0.0002 (0.0024) | 0.0023 (0.0024) | 0.0031 (0.0023) | 0.0034 (0.0023) |
| 2010 | 0.0034 (0.0023) | 0.0001 (0.0023) | 0.0000 (.) | -0.0001 (0.0024) | -0.0019 (0.0024) | 0.0003 (0.0023) | 0.0024 (0.0023) | 0.0032 (0.0023) | 0.0035 (0.0023) |
| 2011 | 0.0036 (0.0024) | 0.0003 (0.0024) | 0.0001 (0.0024) | 0.0000 (.) | -0.0017 (0.0025) | 0.0005 (0.0024) | 0.0025 (0.0024) | 0.0034 (0.0024) | 0.0037 (0.0024) |
| 2012 | 0.0053** (0.0025) | 0.0020 (0.0024) | 0.0019 (0.0024) | 0.0017 (0.0025) | 0.0000 (.) | 0.0022 (0.0024) | 0.0043* (0.0024) | 0.0051** (0.0024) | 0.0054** (0.0024) |
| 2013 | 0.0031 (0.0024) | -0.0002 (0.0024) | -0.0003 (0.0023) | -0.0005 (0.0024) | -0.0022 (0.0024) | 0.0000 (.) | 0.0020 (0.0024) | 0.0029 (0.0024) | 0.0032 (0.0024) |
| 2014 | 0.0011 (0.0024) | -0.0023 (0.0024) | -0.0024 (0.0023) | -0.0025 (0.0024) | -0.0043* (0.0024) | -0.0020 (0.0024) | 0.0000 (.) | 0.0008 (0.0023) | 0.0011 (0.0023) |
| 2015 | 0.0002 (0.0024) | -0.0031 (0.0023) | -0.0032 (0.0023) | -0.0034 (0.0024) | -0.0051** (0.0024) | -0.0029 (0.0024) | -0.0008 (0.0023) | 0.0000 (.) | 0.0003 (0.0023) |
| 2016 | -0.0001 (0.0024) | -0.0034 (0.0023) | -0.0035 (0.0023) | -0.0037 (0.0024) | -0.0054** (0.0024) | -0.0032 (0.0024) | -0.0011 (0.0023) | -0.0003 (0.0023) | 0.0000 (.) |

Standard errors in parenthesis, * p<0.1 ** p<0.05 *** p<0.01

Note: Each box presents the differences in the percentage of household heads without education between two different given years. As the grey becomes darker, the difference on the percentage of household heads without education between the two given years becomes more statistically significant.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.3. Differences in the Education Level of the Household Head - Primary (2008-2016)

| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|-------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 2008 | 0.0000 (.) | -0.0055 (0.0048) | 0.0019 (0.0048) | 0.0059 (0.0047) | 0.0109** (0.0047) | 0.0189*** (0.0047) | 0.0226*** (0.0047) | 0.0223*** (0.0047) | 0.0285*** (0.0046) |
| 2009 | 0.0055 (0.0048) | 0.0000 (.) | 0.0074 (0.0047) | 0.0114** (0.0045) | 0.0164*** (0.0046) | 0.0244*** (0.0046) | 0.0282*** (0.0046) | 0.0279*** (0.0045) | 0.0340*** (0.0044) |
| 2010 | -0.0019 (0.0048) | -0.0074 (0.0047) | 0.0000 (.) | 0.0040 (0.0046) | 0.0090** (0.0046) | 0.0170*** (0.0046) | 0.0208*** (0.0046) | 0.0205*** (0.0045) | 0.0266*** (0.0045) |
| 2011 | -0.0059 (0.0047) | -0.0114** (0.0045) | -0.0040 (0.0046) | 0.0000 (.) | 0.0050 (0.0044) | 0.0130*** (0.0044) | 0.0168*** (0.0045) | 0.0165*** (0.0044) | 0.0226*** (0.0043) |
| 2012 | -0.0109** (0.0047) | -0.0164*** (0.0046) | -0.0090** (0.0046) | -0.0050 (0.0044) | 0.0000 (.) | 0.0080* (0.0045) | 0.0118*** (0.0045) | 0.0115*** (0.0044) | 0.0176*** (0.0044) |
| 2013 | -0.0189*** (0.0047) | -0.0244*** (0.0046) | -0.0170*** (0.0046) | -0.0130*** (0.0044) | -0.0080* (0.0045) | 0.0000 (.) | 0.0037 (0.0045) | 0.0034 (0.0044) | 0.0096** (0.0044) |
| 2014 | -0.0226*** (0.0047) | -0.0282*** (0.0046) | -0.0208*** (0.0046) | -0.0168*** (0.0045) | -0.0118*** (0.0045) | -0.0037 (0.0045) | 0.0000 (.) | -0.0003 (0.0044) | 0.0059 (0.0044) |
| 2015 | -0.0223*** (0.0047) | -0.0279*** (0.0045) | -0.0205*** (0.0045) | -0.0165*** (0.0044) | -0.0115*** (0.0044) | -0.0034 (0.0044) | 0.0003 (0.0044) | 0.0000 (.) | 0.0062 (0.0043) |
| 2016 | -0.0285*** (0.0046) | -0.0340*** (0.0044) | -0.0266*** (0.0045) | -0.0226*** (0.0043) | -0.0176*** (0.0044) | -0.0096** (0.0044) | -0.0059 (0.0044) | -0.0062 (0.0043) | 0.0000 (.) |

Standard errors in parenthesis, * p<0.1 ** p<0.05 *** p<0.01

Note: Each box presents the differences in the percentage of household heads with primary education between two different given years. One of the years is reported in the row, and the other one in the column. As the grey becomes darker, the difference on the percentage of household heads with primary education between the two given years becomes more statistically significant.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.4. Differences in the Education Level of the Household Head - Middle School (2008-2016)

| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|-------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 2008 | 0.0000 (.) | -0.0001 (0.0028) | 0.0036 (0.0028) | 0.0060** (0.0028) | 0.0075*** (0.0028) | 0.0103*** (0.0028) | 0.0120*** (0.0028) | 0.0158*** (0.0028) | 0.0170*** (0.0028) |
| 2009 | 0.0001 (0.0028) | 0.0000 (.) | 0.0037 (0.0026) | 0.0061** (0.0026) | 0.0076*** (0.0026) | 0.0105*** (0.0026) | 0.0121*** (0.0026) | 0.0159*** (0.0026) | 0.0171*** (0.0026) |
| 2010 | -0.0036 (0.0028) | -0.0037 (0.0026) | 0.0000 (.) | 0.0024 (0.0026) | 0.0039 (0.0026) | 0.0068*** (0.0026) | 0.0084*** (0.0026) | 0.0122*** (0.0026) | 0.0134*** (0.0026) |
| 2011 | -0.0060** (0.0028) | -0.0061** (0.0026) | -0.0024 (0.0026) | 0.0000 (.) | 0.0015 (0.0026) | 0.0044* (0.0026) | 0.0060** (0.0026) | 0.0098*** (0.0026) | 0.0110*** (0.0026) |
| 2012 | -0.0075*** (0.0028) | -0.0076*** (0.0026) | -0.0039 (0.0026) | -0.0015 (0.0026) | 0.0000 (.) | 0.0029 (0.0026) | 0.0045* (0.0026) | 0.0083*** (0.0026) | 0.0095*** (0.0026) |
| 2013 | -0.0103*** (0.0028) | -0.0105*** (0.0026) | -0.0068*** (0.0026) | -0.0044* (0.0026) | -0.0029 (0.0026) | 0.0000 (.) | 0.0016 (0.0026) | 0.0054** (0.0026) | 0.0067** (0.0026) |
| 2014 | -0.0120*** (0.0028) | -0.0121*** (0.0026) | -0.0084*** (0.0026) | -0.0060** (0.0026) | -0.0045* (0.0026) | -0.0016 (0.0026) | 0.0000 (.) | 0.0038 (0.0026) | 0.0050* (0.0026) |
| 2015 | -0.0158*** (0.0028) | -0.0159*** (0.0026) | -0.0122*** (0.0026) | -0.0098*** (0.0026) | -0.0083*** (0.0026) | -0.0054** (0.0026) | -0.0038 (0.0026) | 0.0000 (.) | 0.0012 (0.0026) |
| 2016 | -0.0170*** (0.0028) | -0.0171*** (0.0026) | -0.0134*** (0.0026) | -0.0110*** (0.0026) | -0.0095*** (0.0026) | -0.0067** (0.0026) | -0.0050* (0.0026) | -0.0012 (0.0026) | 0.0000 (.) |

Standard errors in parenthesis, * p<0.1 ** p<0.05 *** p<0.01

Note: Each box presents the differences in the percentage of household heads with middle school between two different given years. One of the years is reported in the row, and the other one in the column. As the grey becomes darker, the difference on the percentage of household heads with middle school between the two given years becomes more statistically significant.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.5. Differences in the Education Level of the Household Head - Secondary (2008-2016)

| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|------------------------|
| 2008 | 0.0000 (.) | 0.0010 (0.0032) | -0.0001 (0.0032) | -0.0059* (0.0032) | -0.0065** (0.0032) | -0.0049 (0.0032) | -0.0056* (0.0031) | -0.0142*** (0.0032) | -0.0192*** (0.0032) |
| 2009 | -0.0010 (0.0032) | 0.0000 (.) | -0.0011 (0.0030) | -0.0069** (0.0030) | -0.0075** (0.0031) | -0.0059** (0.0030) | -0.0066** (0.0030) | -0.0152*** (0.0030) | -0.0201*** (0.0030) |
| 2010 | 0.0001 (0.0032) | 0.0011 (0.0030) | 0.0000 (.) | -0.0058* (0.0030) | -0.0064** (0.0031) | -0.0048 (0.0030) | -0.0055* (0.0030) | -0.0141*** (0.0030) | -0.0191*** (0.0030) |
| 2011 | 0.0059* (0.0032) | 0.0069** (0.0030) | 0.0058* (0.0030) | 0.0000 (.) | -0.0006 (0.0031) | 0.0010 (0.0030) | 0.0003 (0.0030) | -0.0083*** (0.0031) | -0.0132*** (0.0030) |
| 2012 | 0.0065** (0.0032) | 0.0075** (0.0031) | 0.0064** (0.0031) | 0.0006 (0.0031) | 0.0000 (.) | 0.0016 (0.0030) | 0.0009 (0.0030) | -0.0077** (0.0031) | -0.0126*** (0.0031) |
| 2013 | 0.0049 (0.0032) | 0.0059** (0.0030) | 0.0048 (0.0030) | -0.0010 (0.0030) | -0.0016 (0.0030) | 0.0000 (.) | -0.0007 (0.0030) | -0.0093*** (0.0030) | -0.0142*** (0.0030) |
| 2014 | 0.0056* (0.0031) | 0.0066** (0.0030) | 0.0055* (0.0030) | -0.0003 (0.0030) | -0.0009 (0.0030) | 0.0007 (0.0030) | 0.0000 (.) | -0.0086*** (0.0030) | -0.0136*** (0.0030) |
| 2015 | 0.0142*** (0.0032) | 0.0152*** (0.0030) | 0.0141*** (0.0030) | 0.0083*** (0.0031) | 0.0077** (0.0031) | 0.0093*** (0.0030) | 0.0086*** (0.0030) | 0.0000 (.) | -0.0050 (0.0030) |
| 2016 | 0.0192*** (0.0032) | 0.0201*** (0.0030) | 0.0191*** (0.0030) | 0.0132*** (0.0030) | 0.0126*** (0.0031) | 0.0142*** (0.0030) | 0.0136*** (0.0030) | 0.0050 (0.0030) | 0.0000 (.) |

Standard errors in parenthesis, * p<0.1 ** p<0.05 *** p<0.01

Note: Each box presents the differences in the percentage of household heads with secondary education between two different given years. One of the years is reported in the row, and the other one in the column. As the grey becomes darker, the difference on the percentage of household heads with secondary education between the two given years becomes more statistically significant.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.6. Differences in the Education Level of the Household Head – Tertiary (2008-2016)

| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|------|-----------------------|-----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 2008 | 0.0000 (.) | 0.0080** (0.0036) | -0.0019 (0.0037) | -0.0023 (0.0036) | -0.0065* (0.0036) | -0.0212*** (0.0039) | -0.0280*** (0.0039) | -0.0237*** (0.0037) | -0.0264*** (0.0036) |
| 2009 | -0.0080** (0.0036) | 0.0000 (.) | -0.0099*** (0.0037) | -0.0103*** (0.0037) | -0.0145*** (0.0036) | -0.0292*** (0.0039) | -0.0359*** (0.0040) | -0.0317*** (0.0037) | -0.0344*** (0.0036) |
| 2010 | 0.0019 (0.0037) | 0.0099*** (0.0037) | 0.0000 (.) | -0.0004 (0.0038) | -0.0046 (0.0037) | -0.0194*** (0.0040) | -0.0261*** (0.0041) | -0.0218*** (0.0038) | -0.0245*** (0.0038) |
| 2011 | 0.0023 (0.0036) | 0.0103*** (0.0037) | 0.0004 (0.0038) | 0.0000 (.) | -0.0042 (0.0037) | -0.0189*** (0.0039) | -0.0256*** (0.0040) | -0.0214*** (0.0038) | -0.0241*** (0.0037) |
| 2012 | 0.0065* (0.0036) | 0.0145*** (0.0036) | 0.0046 (0.0037) | 0.0042 (0.0037) | 0.0000 (.) | -0.0148*** (0.0039) | -0.0215*** (0.0040) | -0.0172*** (0.0037) | -0.0199*** (0.0037) |
| 2013 | 0.0212*** (0.0039) | 0.0292*** (0.0039) | 0.0194*** (0.0040) | 0.0189*** (0.0039) | 0.0148*** (0.0039) | 0.0000 (.) | -0.0067 (0.0043) | -0.0025 (0.0040) | -0.0052 (0.0039) |
| 2014 | 0.0280*** (0.0039) | 0.0359*** (0.0040) | 0.0261*** (0.0041) | 0.0256*** (0.0040) | 0.0215*** (0.0040) | 0.0067 (0.0043) | 0.0000 (.) | 0.0042 (0.0040) | 0.0015 (0.0040) |
| 2015 | 0.0237*** (0.0037) | 0.0317*** (0.0037) | 0.0218*** (0.0038) | 0.0214*** (0.0038) | 0.0172*** (0.0037) | 0.0025 (0.0040) | -0.0042 (0.0040) | 0.0000 (.) | -0.0027 (0.0037) |
| 2016 | 0.0264*** (0.0036) | 0.0344*** (0.0036) | 0.0245*** (0.0038) | 0.0241*** (0.0037) | 0.0199*** (0.0037) | 0.0052 (0.0039) | -0.0015 (0.0040) | 0.0027 (0.0037) | 0.0000 (.) |

Standard errors in parenthesis, * p<0.1 ** p<0.05 *** p<0.01

Note: Each box presents the differences in the percentage of household heads with tertiary education between two different given years. One of the years is reported in the row, and the other one in the column. As the grey becomes darker, the difference on the percentage of household heads with tertiary education between the two given years becomes more statistically significant.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.7. Log income Estimates based on Specification 1
Colombia 2008, 2010, 2012, 2014 and 2016

| Dependent Variable: | 2008 | 2010 | 2012 | 2014 | 2016 |
|---------------------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|
| Log Per Capita Income | (1) | (2) | (3) | (4) | (5) |
| Male | 0.328*** (0.0557) | 0.306*** (0.0421) | 0.213*** (0.0409) | 0.212*** (0.0418) | 0.255*** (0.0422) |
| Household head education | | | | | |
| Primary education | 0.496*** (0.0494) | 0.477*** (0.0328) | 0.440*** (0.0352) | 0.422*** (0.0359) | 0.399*** (0.0372) |
| Middle school | 0.789*** (0.0519) | 0.701*** (0.0381) | 0.694*** (0.0377) | 0.727*** (0.0383) | 0.651*** (0.0392) |
| Secondary education | 1.059*** (0.0532) | 1.026*** (0.0349) | 0.919*** (0.0359) | 0.890*** (0.0383) | 0.858*** (0.0379) |
| Tertiary education | 2.048*** (0.0529) | 1.918*** (0.0386) | 1.770*** (0.0378) | 1.708*** (0.0416) | 1.595*** (0.0411) |
| Male x Primary education | -0.150* (0.0594) | -0.136** (0.0447) | -0.0674 (0.0434) | -0.0508 (0.0446) | -0.0953* (0.0451) |
| Male x Middle school | -0.129* (0.0618) | -0.107* (0.0506) | -0.0465 (0.0462) | -0.117* (0.0490) | -0.140** (0.0484) |
| Male x Secondary education | -0.131* (0.0627) | -0.169*** (0.0465) | -0.0191 (0.0443) | -0.0294 (0.0475) | -0.109* (0.0455) |
| Male x Tertiary education | -0.273*** (0.0620) | -0.190*** (0.0484) | -0.0680 (0.0457) | -0.0301 (0.0495) | -0.132** (0.0480) |
| Constant | 11.83*** (0.0665) | 12.03*** (0.0423) | 12.13*** (0.0511) | 12.30*** (0.0427) | 12.29*** (0.0465) |
| Observations | 162555 | 167596 | 166006 | 162405 | 159879 |
| R-squared | 0.178 | 0.209 | 0.204 | 0.186 | 0.161 |
| Adjusted R-squared | 0.177 | 0.209 | 0.203 | 0.186 | 0.161 |

Clustered errors in parenthesis. *p<0.1 **p<0.05 ***p<0.01

Note: Birth year fixed effects included in all regressions. Results are constrained to the sample of households whose heads were born between 1948 and 1973

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.8. Probability of being poor based on different specifications
Colombia 2008, 2010, 2012, 2014 and 2014

| Dependent variable: Poor status | Specification 1 | | | | Specification 2 | | | |
|------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | 2008 (1) | 2010 (2) | 2012 (3) | 2014 (4) | 2008 (5) | 2010 (6) | 2012 (7) | 2014 (8) |
| Male | -0.2382*** (0.0562) | -0.0558 (0.0565) | -0.1246** (0.0551) | -0.1948*** (0.0558) | -0.2595*** (0.0585) | -0.1102* (0.0582) | -0.1737*** (0.0585) | -0.2256*** (0.0599) |
| Primary Education | -0.6477*** (0.0487) | -0.4969*** (0.0507) | -0.5389*** (0.0468) | -0.5408*** (0.0483) | -0.5999*** (0.0497) | -0.4582*** (0.0518) | -0.5020*** (0.0478) | -0.5069*** (0.0502) |
| Middle School | -0.9493*** (0.0520) | -0.8604*** (0.0538) | -0.9113*** (0.0507) | -0.8381*** (0.0532) | -0.8677*** (0.0530) | -0.7918*** (0.0549) | -0.8261*** (0.0519) | -0.7680*** (0.0552) |
| Secondary Education | -1.3215*** (0.0519) | -1.0930*** (0.0527) | -1.1295*** (0.0499) | -1.0561*** (0.0519) | -1.1622*** (0.0530) | -0.9416*** (0.0545) | -0.9873*** (0.0507) | -0.9234*** (0.0537) |
| Tertiary Education | -2.1906*** (0.0562) | -1.9381*** (0.0609) | -1.8836*** (0.0574) | -1.7338*** (0.0565) | -1.8611*** (0.0591) | -1.6394*** (0.0643) | -1.6006*** (0.0593) | -1.4407*** (0.0592) |
| Male x Primary Education | 0.1142* (0.0605) | -0.1055* (0.0610) | -0.0028 (0.0601) | 0.0703 (0.0605) | 0.1620*** (0.0618) | -0.0267 (0.0622) | 0.0567 (0.0619) | 0.1154* (0.0628) |
| Male x Middle School | 0.0592 (0.0629) | -0.0987 (0.0657) | 0.0058 (0.0649) | 0.0258 (0.0684) | 0.1762*** (0.0648) | 0.0673 (0.0671) | 0.1278* (0.0676) | 0.1176* (0.0709) |
| Male x Secondary Education | 0.0735 (0.0624) | -0.2345*** (0.0630) | -0.1158* (0.0621) | -0.0770 (0.0642) | 0.2298*** (0.0646) | -0.0357 (0.0644) | 0.0657 (0.0651) | 0.0668 (0.0678) |
| Male x Tertiary Education | 0.0599 (0.0676) | -0.2161*** (0.0723) | -0.0755 (0.0727) | -0.0137 (0.0703) | 0.1470** (0.0705) | -0.0449 (0.0753) | 0.0864 (0.0762) | 0.0663 (0.0739) |
| Year of birth fixed effect | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Labor market controls | | | | | ✓ | ✓ | ✓ | ✓ |
| Household Characteristics | | | | | | | | |
| Exposure to shocks | | | | | | | | |

Standard errors in parenthesis. *p<0.1 **p<0.05 ***p<0.01

Note: Results are constrained to the sample of households whose heads were born between 1948 and 1973. Labor market controls included the employment status, sector of the economy and type of employment. Household characteristics includes household size and access to basic services as water, electricity and sewage. Exposure to shocks includes losing the job and the cause.

Source: Own estimations based on GEIH from 2008 to 2016.

Table C.8. (continued)

| Dependent variable: Poor status | Specification 3 | | | | Specification 4 | | | |
|------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | 2008 (9) | 2010 (10) | 2012 (11) | 2014 (12) | 2008 (13) | 2010 (14) | 2012 (15) | 2014 (16) |
| Male | -0.2017*** (0.0611) | -0.0255 (0.0662) | -0.1058* (0.0627) | -0.1589** (0.0636) | -0.2877*** (0.0604) | -0.0763 (0.0663) | -0.1502** (0.0629) | -0.2204*** (0.0632) |
| Primary Education | -0.4046*** (0.0537) | -0.2119*** (0.0597) | -0.2816*** (0.0523) | -0.3369*** (0.0527) | -0.4231*** (0.0527) | -0.2166*** (0.0596) | -0.2884*** (0.0520) | -0.3527*** (0.0522) |
| Middle School | -0.5211*** (0.0580) | -0.4151*** (0.0632) | -0.5145*** (0.0565) | -0.4837*** (0.0591) | -0.5515*** (0.0570) | -0.4284*** (0.0631) | -0.5260*** (0.0563) | -0.5070*** (0.0587) |
| Secondary Education | -0.6815*** (0.0572) | -0.4752*** (0.0623) | -0.5484*** (0.0555) | -0.5130*** (0.0562) | -0.7187*** (0.0562) | -0.4914*** (0.0622) | -0.5660*** (0.0555) | -0.5416*** (0.0558) |
| Tertiary Education | -1.1738*** (0.0629) | -0.9848*** (0.0726) | -1.0286*** (0.0634) | -0.8800*** (0.0612) | -1.2232*** (0.0621) | -1.0136*** (0.0729) | -1.0535*** (0.0638) | -0.9275*** (0.0609) |
| Male x Primary Education | 0.1208* (0.0653) | -0.1116 (0.0699) | -0.0050 (0.0661) | 0.1188* (0.0660) | 0.1475** (0.0644) | -0.1049 (0.0699) | 0.0052 (0.0660) | 0.1392** (0.0656) |
| Male x Middle School | 0.0858 (0.0688) | -0.0424 (0.0746) | 0.0727 (0.0706) | 0.0826 (0.0750) | 0.1251* (0.0680) | -0.0258 (0.0745) | 0.0886 (0.0705) | 0.1086 (0.0746) |
| Male x Secondary Education | 0.1003 (0.0674) | -0.1400* (0.0722) | -0.0321 (0.0691) | -0.0129 (0.0702) | 0.1527** (0.0666) | -0.1186 (0.0721) | -0.0072 (0.0692) | 0.0207 (0.0700) |
| Male x Tertiary Education | 0.0048 (0.0732) | -0.1573* (0.0833) | -0.0003 (0.0796) | -0.0404 (0.0762) | 0.0626 (0.0725) | -0.1279 (0.0835) | 0.0299 (0.0801) | 0.0028 (0.0761) |
| Year of birth fixed effect | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Labor market controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Household Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Exposure to shocks | | | | | ✓ | ✓ | ✓ | ✓ |

Standard errors in parenthesis. *p<0.1 **p<0.05 ***p<0.01

Note: Results are constrained to the sample of households whose heads were born between 1948 and 1973. Labor market controls included the employment status, sector of the economy and type of employment. Household characteristics includes household size and access to basic services as water, electricity and sewage. Exposure to shocks includes losing the job and the cause.

Source: Own estimations based on GEIH from 2008 to 2016.