A Methodology for Updating International Middle-Class Lines for the Latin American and Caribbean Region

Jaime Fernandez
Sergio Olivieri
Diana Sanchez

WORLD BANK GROUP
Poverty and Equity Global Practice
May 2023

Abstract

The middle class in Latin America and the Caribbean has been a central focus of policy debates in the region since the COVID-19 pandemic began. To identify and track vulnerable and middle-class populations accurately, it is necessary to update the upper and lower bounds for the middle class using 2017 purchasing power parity exchange rates. This paper contributes with a two-step methodology for updating these thresholds. The method indicates that updating the $13 lower-bound line in 2011 purchasing power parity dollars to 2017 purchasing power parity dollars results in a vulnerability line of $14. The study also finds an upper bound of $81 per person per day in 2017 purchasing power parity, compared with $70 in 2011 purchasing power parity. These thresholds are robust to a variety of assumptions and methodologies. The results of this study indicate that the proportion of the population in Latin America and the Caribbean classified as middle class increased from 36.3 percent in 2011 to 37.2 percent in 2017. However, there were no significant changes in the characteristics of this group.

A Methodology for Updating International Middle-Class Lines for the Latin American and Caribbean Region[1]

Jaime Fernandez, Sergio Olivieri, and Diana Sanchez[2]

I.   Introduction

In 2018, for the first time in nearly two decades, Latin America and the Caribbean's middle class became the largest socioeconomic group. It increased from more than a fifth of the LAC population in 2000 (21.6 percent) to more than a third in 2019 (37.6 percent), based on 2011 PPPs. However, during the pandemic, there was a rapid decline in the size of this group in most countries. As a result, LAC is no longer a middle-class region. This group shrunk by four percentage points in 2020, excluding Brazil, representing 13 million people falling into poverty. Moreover, this decrease reached similar levels as those in 2013. Peru, Colombia, and Argentina drove this significant reduction in 2020.[3] Governments must continue targeting policies to support the most vulnerable populations, particularly after the COVID-19 pandemic, followed by the Russian Federation–Ukraine war, which significantly impacted the region. Thus, it is important to accurately measure the size of the vulnerable population, monitor its evolution, and know where they live and their characteristics.

The major challenge regarding estimating the LAC region's vulnerable and middle class is the identification of the lower and upper thresholds defined initially by Lopez-Calva & Ortiz-Juarez (2014) and Ferreira et al. (2013). To do so, the principal data that allow for the comparability of different countries' living standards are purchasing power parities (PPPs). In May 2020, the International Comparison Program (ICP) published new 2017 PPPs. The 2017 PPPs reflect the most recent relative price differences across a wide range of countries around the world. Jolliffe et al. (2022) assessed the impact of the 2017 PPPs on global poverty by updating the three international thresholds: the $1.9 2011 PPP line to $2.15 2017 PPP per person per day, $3.2 2011 PPP to $3.65 2017 PPP per person per day, and $5.5 2011 PPP to $6.85 2017 PPP per person per day.

These changes pose a challenge for how to update the upper and lower boundaries for identifying the vulnerable and middle-class populations in LAC countries. This paper makes three contributions to the literature on global vulnerability and middle-class measurement. First, it proposes a two-step methodology to update these thresholds when new and better rounds of PPPs are estimated. In the past, when 2011 PPPs were published, a simple approach was implemented to update the thresholds. The recent implementation of new PPPs presents an opportunity to revisit and improve previous methodologies used to define middle-class thresholds. However, replicating past approaches would not address the issues with those methodologies. Therefore, this study proposes a new methodology that addresses and resolves some of these problems in some cases and mitigates them in others. Thus, the new lines proposed in this document are determined by i) the use of the revised PPPs, ii) the methodological innovations introduced, and iii) the larger number of countries for which data is available. This novel approach yields a vulnerability or lower-bound line of $14 in 2017 PPP per person per day, compared to $13 in 2011 PPP. The study shows that this rounded value is robust to alternative approaches.

Second, this paper proposes a clear definition of the upper bound. Lopez-Calva & Ortiz-Juarez (2014) and Ferreira et al. (2013) do not provide an equally in-depth discussion of the upper bound as they do for the

---

[3] Including Brazil, the middle class declined to 36 percent of the population in 2020, resulting in the loss of 7 million people from this socioeconomic class regionally.

lower bound since, firstly, moving it up (down) the income distribution includes (excludes) a small percentage of the population; secondly, an income threshold above $50 in 2005 PPP would limit the representativeness of the upper class in some countries; and, thirdly, household surveys do not collect realistic information on the richest population. Using the proposed methodology in this study with the 2017 PPP, the upper bound for the middle class reaches $81 per person per day, compared with $70 in the 2011 PPP. It is shown that this rounded value is robust to different methodologies.

Third, this study analyzes the impact on the Latin American and Caribbean regional poor, vulnerable, and middle-class groups estimated using the international poverty lines (IPL), the regional vulnerability and middle-class lines, and the 2017 PPPs. The 2017 PPPs would slightly increase historical estimates of poverty at the $2.15 and $3.65 lines, but significantly at $6.85, compared with the 2011 PPPs. Extreme poverty (measured at the $2.15 line) and poverty at $3.65 would increase marginally by 0.3 and 0.9 percentage points (pp), respectively in 2017, with the 2017 PPPs. The regional count of the extreme poor increased by 1.9 million, which is largely driven by 1.3 million and 300,000 more poor people in Brazil and the Andean region, while poverty increased in all other subregions except the Southern Cone.[4] Considering the upper-middle-income country line, $6.85 per person per day, the incidence of poverty in the region rose around five percentage points in 2017 with the 2017 PPP. The headcount shifted almost five percentage points since 2000, compared with the 2011 PPP series. This represents 27 million more poor people in LAC, driven by 12 million and 5.5 million in Brazil and the Andean region, respectively.

A significant increase in the poor population is mirrored by a decrease in the size of the vulnerable group by around 5.4 pp in 2017 with the revised lower-middle-class bound of $14 in 2017 PPP. The regional vulnerability trend is due to declines in the Andean region and Central America, offset by increases in countries in the Southern Cone, including Brazil, since 2012. The update of the upper-middle-class threshold of $81 in 2017 PPP increases the size of the middle class regionally by around 0.9 pp. This change in lower and upper middle-class limits provides a similar positive trend for this group over the last two decades.

The rest of the paper is organized as follows. Section II presents the methodological details for updating both the vulnerability and middle-class lines in terms of the 2017 PPPs. Section III describes the data used in the study. Section IV presents the estimation results for the middle-class upper and lower bounds and some sensitivity analysis. Section V documents changes observed in regional and country-level poverty, considering the vulnerable and middle-class estimates with the 2017 PPP. Finally, Section VI discusses the final remarks.

---

[4] The Andean region is the aggregate of Bolivia, Colombia, Ecuador, and Peru; Central America is the aggregate of Costa Rica, Guatemala, Honduras, Nicaragua, Panama, El Salvador, and the Dominican Republic, and the Southern Cone is the aggregate of Argentina, Chile, Paraguay, and Uruguay.

II.   Methodological approach

The study proposes a two-step methodology for updating the LAC region middle-class thresholds. The first step consists of constructing synthetic panels given the lack or scarcity of long panels in most countries. The second step defines the lower and upper middle-class thresholds.

II.1 Constructing synthetic panels:

Ideally, a proper study of welfare dynamics entails following the same observation (household or individual) for at least two—or preferably, multiple—periods. 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 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 (2006) 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 for analyzing 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.

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 a perfect positive correlation between error terms) for the transitions using two rounds of cross sections. In addition, Dang and Lanjouw (2013, 2016) extend this method and calculate point estimates of poverty mobility based on the synthetic panels. However, this approach relies on the key assumption that the residual terms of the income equations in two periods are distributed according to a bivariate normal distribution.

To avoid the strong distribution assumption for residual or error correlation estimates, Lucchetti et al. (2020) introduce a Least Absolute Shrinkage and Selection Operator method with multiple imputations by Predictive Mean Matching (LASSO-PMM) for constructing synthetic panels. Among the several advantages this method has over those above, a very important one is that it does not require estimating any error correlation terms or assuming a certain distribution of residuals in the underlying regressions. As mentioned by Ñopo (2004), matching avoids any parametric assumptions that may impose restrictions on the behavior of the random variables involved in the analysis. The LASSO-PMM method allows for obtaining point estimates as well as upper and lower bounds of welfare dynamics. This penalized regression method uses regularization and takes advantage of machine learning techniques to minimize the mean square error (MSE) of predictions. As a result, it allows for the estimation of more accurate welfare predictions outside the estimation sample compared to traditional regression models, since variables and model selection are performed automatically by penalizing the coefficients and cross-validation (Tibshirani, 1996). Lucchetti et al (2020) found that the LASSO-PMM predictions are statistically indistinguishable from actual poverty rates, transitions, and income changes calculated using

actual panels.[5] On average, the predicted poverty rates using LASSO-PMM were one percentage point away from the observed rates, when using data from the four validation countries.

To build the synthetic panels required to define the new vulnerability and middle-class lines, this paper follows Lucchetti et al. (2020) as a starting point. and introduces several improvements in its methodology. First, matching is performed only between subsets (donation classes) of individuals with the same time-invariant features. These constrained donation classes are built based on gender, birth year, and educational attainment. Second, sampling weights are used to harmonize the joint distributions of identified common variables across the two surveys (Renssen, 1998). This method (known as Renssen's procedure) consists of a series of calibration steps of the survey weights (in both the donor and receiver surveys[6]) implemented to achieve consistency between some aggregate estimates, e.g.: income, gender or age distributions. Third, when several donors are equidistant in terms of the predicted income, one of them is randomly chosen according to their sampling weights. The fourth improvement is a direct consequence of the two previous ones: this method allows to preserve the marginal distribution of income from the donor survey, which is essential not to alter welfare dynamics. In addition, although synthetic panels have their own set of limitations, they represent a significant improvement over using cross-sectional data to estimate the probability of falling into poverty. Without actual panels available, this would have been the case.

Starting from two available rounds (i.e., j = rounds 1 and 2) of cross-sectional microdata. Let y_{ij} be the per capita household income for household i, (i.e., $i = 1, …, N$), in survey round j with sample size N, and X_{ij} be a vector of household head characteristics observed for household i in survey round j. These household head characteristics, observed in both survey years, can include time-invariant variables (e.g., household head's gender, ethnicity, place of birth, etc.), deterministic variables (e.g., such as age, education, literacy, etc.), and retrospective questions collected in round 2 about round 1 such as assets (Cruces et al, 2015; Dang and Lanjouw, 2018). Changes in household composition can help identify matching errors, which could lead to spurious measures of change. To avoid this, the estimation samples are constrained to household heads' age: 25 to 65 in the first cross-section and adjust this age range accordingly in the second cross-section.

The linear model where i-th household's log per capita income[7] y_{ij} is explained solely by household head characteristics X_{ij} for each survey round is given by:

$$y_{ij} = \alpha_j + \beta_j^T X_{ij} + \varepsilon_{ij} \qquad j = 1, 2 \qquad (1)$$

where ε_{ij} is an error term and X_{ij} is a vector of *K* regressors, and α_j is the intercept. In this context, $\alpha_j + \beta_j^T X_{ij}$ would represent the portion of log income explained exclusively by deterministic and time-invariant household head characteristics.

---

[5] Lucchetti et al (2020) use 36 panels from Argentina, Chile, Peru, and Nicaragua to validate their estimates of mobility using harmonized variables frequently used in many regional and global studies. After the validation exercises, the authors implemented the LASSO-PMM method in 43 countries worldwide.

[6] The donor survey is the one that contains the actual income data from the first-round survey. The receiver is the second-round survey where the income is imputed.

[7] To simplify notation, in this document y_{ij} will be used to refer to the logarithm of household per capita income.

The objective is to calculate, for household i interviewed in round 2, the change in log income between the two survey rounds:

$$\Delta y_i = y_{i2} - y_{i1} \qquad\qquad (2)$$

where y_{i1} and y_{i2} are respectively the first and second round log per capita incomes of household i surveyed in round 2. Therefore, the log per capita household income in round 1 for household i is unknown and must be estimated. These changes would be easily calculated with panel data since all households are interviewed in both rounds (i.e., Δy_i is observed for every household i interviewed in both rounds and does not require to be estimated). Given panel data is costly and scarce in most countries in the region, synthetic panels allow for predicting the first round of "unobserved" incomes of households surveyed in the second. As noted above, to avoid making assumptions about the behavior of the residuals, in this approach the unobserved log income is not obtained directly from an econometric prediction, but through statistical matching between household heads from the two surveys. Thus, the predicted income from model (1) is one of the variables on which the closeness between observations is measured. More specifically, y_{i1} and y_{i2} are predicted using LASSO regression as follows:

$$\hat{y}_{ij} = \hat{\alpha}_{LASSO_j} + \hat{\beta}^T_{LASSO_j} X_{ij} \qquad\qquad j = 1,2 \qquad (3)$$

where $(\hat{\alpha}_{LASSO_j}, \hat{\beta}^T_{LASSO_j}) = \mathrm{argmin}_{(\alpha_j,\beta_j)} \left[\sum_{i=1}^{N_j} (y_{ij} - \alpha_j - \beta_j^T X_{ij})^2 + \lambda \|\beta_j\|_1 \right]$, λ is the penalization parameter chosen through cross-validation, and $\|\beta_j\|_1$ is the ℓ_1 norm of the coefficient vector β_j.

This method estimates the portion of log income that is assumed to be time-invariant, at least when using cross-sectional surveys close in time. Although this assumption does not always hold, there are several reasons why it may be reasonable to believe that the portion of income explained by time-invariant covariates remains constant in close periods. First, personal characteristics such as education level, work experience, and skill level tend to be relatively stable over time and may have a persistent effect on income. Second, if individuals have limited mobility and are unable to easily change their geographic location or industry of employment (which is expected to be the case in close periods), this could limit their ability to access higher-paying jobs and result in a relatively constant portion of income being explained by time-invariant factors. Third, if economic conditions (such as the overall level of demand for goods and services) remain relatively constant over time, this could also lead to the portion of income explained by time-invariant factors remaining constant.

If a purely parametric approach were used, the full welfare measure in round 1 would be completed by adding some residuals to $\hat{y}_{ij}$. That is not the case in the method presented in this paper. The main assumption behind this semi-parametric approach is that the portion of log income explained by time-invariant covariates remains constant in both periods. As a result, $\hat{y}_{i1}$ and $\hat{y}_{i2}$ can be used to find the most similar household heads in both surveys (within the set of observations that share common characteristics) and thus build the synthetic panel. For every observation in the second round of data, a set of neighbors (in terms of $\hat{y}_{i1}$ and $\hat{y}_{i2}$ closeness) is found in the first-round data by looking for observations with the smallest absolute difference between the two linear predictions. Among these neighbors, one of them is randomly selected based on the sampling weights from the first round. In addition, the matching method requires that the closest neighbors are searched only within a pre-defined

subset of households' heads having the same gender, educational attainment and age. The observed income from that neighbor is then imputed to the corresponding observation in the second round, which represents the first-round log income for that household surveyed in the second round:

$$\tilde{\Delta}y_i = y_{i2} - \tilde{y}_{i1} \qquad (4)$$

where $\tilde{y}_{i1}$ is actual first-round observed income for the household chosen through the matching process described above.

Summarizing, the steps to estimate the "unobserved" household per capita income in the first round are the following:

1. Append first and second round of cross-sectional data and create a fused dataset
2. Harmonize the joint distribution of gender and educational attainment in both surveys
3. Take a sample for the statistical learning stage: 80 percent of the fused dataset defined in the first step
4. Estimate the parameters and select the best lambda through cross-validation
5. Calculate the LASSO linear prediction of the first-round log incomes for all households surveyed in the first round as in equation (3)
6. Obtain the LASSO linear fit of the second-round log incomes for all households surveyed in the second round as in equation (3)
7. For every household head m in the second round, obtain the nearest first-round neighbors by minimizing the absolute difference between $\hat{y}_{m2}$ and $\hat{y}_{i1}$ for every observation i in the first round that has the same gender, educational attainment and birth year
8. Randomly select one neighbor s from the list of nearest neighbors based on first-round sampling weights
9. Take the observed log income of neighbor s (y_{s1}) chosen in the first round, and assign it to observation m surveyed in the second round ($\tilde{y}_{m1} = y_{s1}$)
10. Estimate movements in-out of poverty and other income dynamics
11. Repeat 100 times all steps 3 to 10

This process allows for estimating standard errors for the statistics in interest. For every bootstrap sample of the data, a new LASSO is fit and the whole welfare vector is constructed via the improved LASSO PMM process described above. As a result, both point and interval estimates are calculated for every poverty dynamics indicator.

II.2 Defining the middle-class thresholds

The major challenge regarding the estimation of the LAC region's vulnerable and middle-class groups is the identification of the lower and upper thresholds. These were defined initially by Lopez-Calva & Ortiz-Juarez (2014) and Ferreira et al. (2013). These authors focus mainly on defining and computing the lower bound. An individual is defined as vulnerable if the probability of falling back into poverty over a five-year interval is greater than 10 percent. Both references use panel data from three different sets of countries. While Lopez-Calva & Ortiz-Juarez (2014) use Chile, Mexico, and Peru, Ferreira et al. (2013) use Argentina,

Colombia, and Costa Rica. Both reach a similar lower bound estimate of $10 per person per day in 2005 PPPs.

The middle-class upper threshold does not follow a thorough definition or computation as the lower bound. Birdsall et al. (2011) show that varying the upper threshold from $50 to $100 a day would move the percentile of the LAC's elite from the top 2.2 percent to the top 0.5 percent. Thus, both define the upper threshold as $50 per person per day in 2005 PPP which is equivalent to the top 2.2 percent of the population.

More recently, when 2011 PPPs were published, a very simple approach was implemented without applying previous methodologies.[8] Firstly, using each country's 2005 PPP conversion factor, the vulnerable and middle-class lines were converted to local currency units at 2005 prices. Secondly, these values were deflated to 2011 prices using each country's Consumer Price Index (CPI) and converted back to US dollars using their corresponding 2011 PPP conversion factors. Finally, a simple average of the resulting lines was estimated to obtain a regional value. By rounding to the closest unit, the lower and upper middle-class lines in 2011 PPP for LAC were then set at $13 and $70 a day, respectively.

The proposed approach estimates the LAC lower middle-class bound (vulnerability line) in two steps. First, it builds a two-year synthetic panel for each country using every pair of available cross-section years between 2010 and 2019. Then, the vulnerability line is defined as the median per capita household income over households who were not poor in the initial year and became poor in the final year, i.e., a household i in $j = 2$ moves into poverty if $y_{i2} < z$ and $\tilde{y}_{i1} > z$, where z is the international poverty line of 6.85 a day in purchasing power parities (PPP) 2017.

There are a few reasons why it might be useful to define the lower bound of the middle class as the median of the income distribution of people who have recently moved out of poverty. First, the median income represents the midpoint of the income distribution, so it is less affected by extreme values that could skew the results. This can help to ensure that the middle-class thresholds are fair and representative of the income needs of a typical household.

A second advantage of using the median income as the lower bound of the middle class (upper bound of the vulnerable group) is that it highlights the fact that the vulnerable population is not static, but rather constantly changing. Additionally, using a more reliable statistic like the median allows for a more precise estimate of the minimum income needed to be considered middle class. It is important to note that it does not imply that the probability-based approach (Lopez-Calva & Ortiz-Juarez, 2014) fails to fulfill these conditions; both methods acknowledge the dynamic nature of the vulnerable population.

Finally, employing the median income as the lower bound of the middle class can contribute to making the middle-class thresholds more responsive to economic changes and shifts in income distribution. As the economy grows and more people rise out of poverty, the middle-class thresholds can be adjusted accordingly to reflect these changes. This responsiveness allows for a more inclusive definition of both the vulnerable and middle-class populations, preventing the exclusion of individuals who are making

---

[8] World Bank (2021).

progress yet may still be considered poor or vulnerable by current standards. Formally, the vulnerability line for each country in a given two-year synthetic panel (vl) is defined such that:

$$F_{Y_1}(vl) = 0.5 \qquad\qquad (5)$$

where $F_{Y_1}(.)$ is the empirical cumulative distribution function of per capita income in year 1, restricted to the subset of people who went from not being poor in year 1 to being poor in year 2.

This estimation process is repeated 100 times for each two-year synthetic panel, which generates a distribution of vulnerability lines for each country.[9] This allows for both point estimates and confidence intervals for the lower middle-class bound. Finally, to estimate the line at the regional level, the method averages out all these medians across countries and over time.[10]

To draw the upper-middle-class threshold (middle-class line), the study follows the same methodological approach presented for the lower-middle-class bound. Conceptually, it is proposed to define the middle-class line as the maximum per capita income of households that went from being non-poor to being poor in each two-year synthetic panel. However, because this measure can be very volatile, the 99th percentile of the income distribution of those who went from not being poor in year 1 to being poor in year 2 is used instead.

It has been suggested that the boundaries for the middle class should be defined using quantiles because the factors that influence the likelihood of moving from non-poverty to poverty may vary between countries. Otherwise, the specific criteria used to determine the boundaries would depend on the model being used and the data available. Thus, using a consistent definition across countries, it becomes easier to compare the size, stability, and characteristics of the middle class in different Latin American countries. This can help to identify patterns and trends and inform regional policy development.

Formally, the middle-class line for each country in a given two-year synthetic panel (mcl) is defined such that:

$$F_{Y_1}(mcl) = 0.99 \qquad\qquad (6)$$

where $F_{Y_1}(.)$ is once again the empirical cumulative distribution function of per capita income of those who experienced downward mobility at the 6.85 line (PPP 2017). Lastly, it averages these statistics across countries and over time after repeating the process 100 times for each country and each synthetic panel. Thus, point estimates and confidence intervals are obtained for the upper-middle-class threshold.

---

[9] The approach considers the median instead of the mean due to not being affected by extreme values and being consistent with the international poverty threshold methodology (see Jolliffe and Prydz, 2016).

[10] It is relevant to point out that results do not significantly vary when weighted averages based on population are considered.

III.   Data

This paper uses harmonized cross-section microdata for 15 countries in Latin America and the Caribbean (i.e., Argentina, Bolivia, Brazil, Colombia, Costa Rica, Chile, Dominican Republic, El Salvador, Ecuador, Honduras, Mexico, Panama, Paraguay, Peru, and Uruguay) for the series starting in 2010 and ending in 2019.[11] These data are from the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), a joint effort of the World Bank and the Center for Distributive, Labor, and Social Studies (CEDLAS) at the National University of La Plata in Argentina.[12] The selection of these countries is based on the availability of at least two comparable data points between 2010 and 2019 within a two-year interval and on the accessibility of the necessary variables to conduct the estimation. More than 85 synthetic panels were constructed for the selected countries with a two-year length following Balcazar et al (2018). Table 1 summarizes the countries, initial and end years, household-survey names in SEDLAC, and the total number of synthetic panels. This study considers information for small and large economies in the region; lower-middle, upper-middle, and high-income countries, and represents 94 percent of the LAC population. The countries excluded from the exercise are those whose data is non-existent or not available circa the end year of the interval.

Table 1: Available cross-section surveys

Country	Years	Survey	Total # SP	Classification
Argentina	2010,2011, 2012, 2013, 2014,2016, 2017, 2018, 2019	Encuesta Permanente de Hogares- Continua	6	H
Bolivia	2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta Continua de Hogares	7	LM
Brazil	2012,2013, 2014, 2015, 2016, 2017, 2018,2019	Pesquisa Nacional por Amostra de Domicilios	6	UM
Colombia	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Gran Encuesta Integrada de Hogares	8	UM
Costa Rica	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta Nacional de Hogares	8	UM
Chile	2006, 2009, 2011, 2013, 2015, 2017	Encuesta de Caracterización Socioeconómica Nacional	3	H
Dominican Republic	2017,2018, 2019	Encuesta de Fuerza de Trabajo	1	UM
Ecuador	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta de Empleo, Desempleo y Subempleo	8	UM
El Salvador	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta de Hogares de Propósitos Múltiples	7	LM

---

[11] Guatemala, Nicaragua, and Haiti were not included in the analysis due to the lack of microdata circa 2019.
[12] The SEDLAC project consists of more than 400 household surveys in more than 25 LAC countries to provide statistics on poverty and other distributional and social variables. See Bourguignon (2015) for a detailed description of the SEDLAC project.

Honduras	2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta Permanente de Hogares Propósitos Múltiples	7	LM
Mexico	2016, 2018, 2020	Encuesta Nacional de Ingresos y Gastos de Hogares	1	UM
Panama	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta de Hogares	8	H
Paraguay	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta Permanente de Hogares	7	UM
Peru	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta Nacional de Hogares	8	UM
Uruguay	2010,2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019	Encuesta Continua de Hogares	7	H

Note: LM: Lower-Middle Income, UM: Upper-Middle Income, and H: High Income
Source: SEDLAC (CEDLAS and World Bank) for country classification:
https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

The proposed approach estimates a model of household-level income using a set of time-invariant and deterministic harmonized variables as controls to provide estimates of welfare dynamics. The harmonization procedure requires cleaning and processing each cross-section to ensure that all definitions and variables are identical in each country and year. The full list of available covariates for each country's data is shown in Table 2.

Table 2: Harmonized variables for LASSO-PMM

Variable	Definition
Welfare aggregate	Per household capita income for international poverty estimations. It is expressed daily basis and deflated using 2017 PPPs
Age	Age of the household head (level and squared)
Gender	Gender of the household head
Education Level	Education of the household head in four education levels: (i) No education (ii) Primary (complete or incomplete) (iii) Secondary (complete or incomplete) (iv) Tertiary (complete or incomplete)
Literacy	Literacy status of household head
Weight	Survey weights

Source: SEDLAC (CEDLAS and World Bank)

When estimating regional middle-class thresholds, income distributions are converted into a common, internationally comparable currency unit using exchange rates. PPP conversion factors are preferred to market exchange rates since it incorporates both the relative prices of tradable goods as the market exchange rates and non-tradable services (e.g., getting a haircut) across countries (Jolliffe et al. 2022). PPPs measure how much it costs to purchase a basket of goods and services in one country compared to how much it costs to purchase the same basket of goods and services in terms of the currency units of a reference country, typically the United States. The estimates of PPP and market exchange rates are from the ICP, and the study uses the PPP estimates from the 2011 and 2017 rounds.

IV.   Derivation of middle-class thresholds

The regional lower and upper-middle-class thresholds have been previously derived without following the Lopez-Calva & Ortiz-Juarez (2014) or Ferreira et al. (2013) methodology. This study proposes a method that could be replicated in future PPPs rounds with the following advantages. First, synthetic panel data provides a robust methodological alternative when panel data is scarce or nonexistent and allows increasing the sample from three to 15 countries which might induce greater statistical support. Including all types of countries in the LAC region (i.e., lower-middle, upper-middle, and high-income countries) also contributes to the robustness of the measure. Second, using an updated series of household surveys during the last decade (i.e., 2010 to 2019) represents current patterns and avoids methodological changes introduced by National Statistics Offices, weakening comparisons of countries over time and across countries.[13] Third, the paper proposes the median per capita income of those who changed their poverty status (i.e., for the lower-bound), which makes the final threshold less vulnerable to outliers and aligned with the IPL methodology. Finally, another advantage is that results are not sensitive to model specification as in the Lopez-Calva and Ortiz-Juarez (2014) method.

The results of using the methodology proposed to update the middle-class thresholds are shown below. Regional lower and upper bounds for measuring the middle class are robust to choosing a broader set of countries, defining a different time interval, and implementing different methodologies. Based on the proposed approach, the median lower-middle-class bound, or vulnerability line is estimated to be $14 per person per day (2017 PPP). The upper-middle-class bound, or middle-class line is calculated to be $81 per person per day (2017 PPP). These results were obtained using 38 synthetic panels built with available surveys within +/- 2 years circa 2017. Figure 1 illustrates both thresholds for the LAC region.

Figure 1: Middle-class lower and upper bound for LAC countries in 2017 PPP
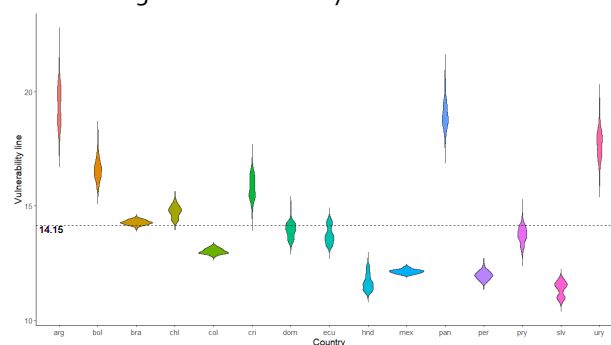
Panel A: Regional vulnerability line   Panel B: Regional middle-class lines
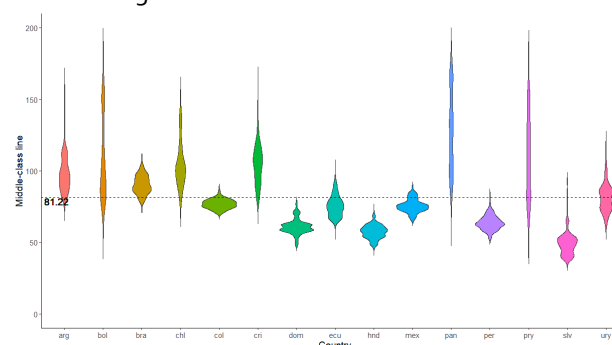
Note: The amplitude of the empirical density presented in the violin plot for each country is directly associated with the size of the confidence interval for each line, based on 100 bootstrap repetitions
Source: Own estimations based on SEDLAC (2022)

---

[13] For instance, in the Mexican case only the 2016 and 2018 synthetic panel was constructed because household surveys are not directly comparable with the historical series. Therefore, the analysis derived from these data should not be compared with pre-2016 numbers.

Robustness check: Expanding the set of countries

Lopez-Calva and Ortiz-Juarez (2014) and Ferreira et al. (2013) only use a particular set of high and upper-middle-income countries in LAC to define the regional middle-class bounds. The limitation of the selection of countries depends on the availability of longitudinal data. Given that this is not a restriction in the proposed approach, expanding the number of countries, including high, upper-middle, and low-middle income, enhances the estimation results by increasing regional representativeness. Following Jolliffe et al. (2022) to check the robustness of results, both lower and upper middle-class bounds were calculated cumulatively by ranking countries from lowest to highest GDP per capita. Figure 2 shows that both vulnerability and middle-class lines estimates are robust to using fewer countries. Each point in the figure corresponds to the line estimated using the synthetic panels available for that country and all those to the left.

Figure 2: Cumulative middle-class bounds for LAC countries in 2017 PPP ordered by country GDP
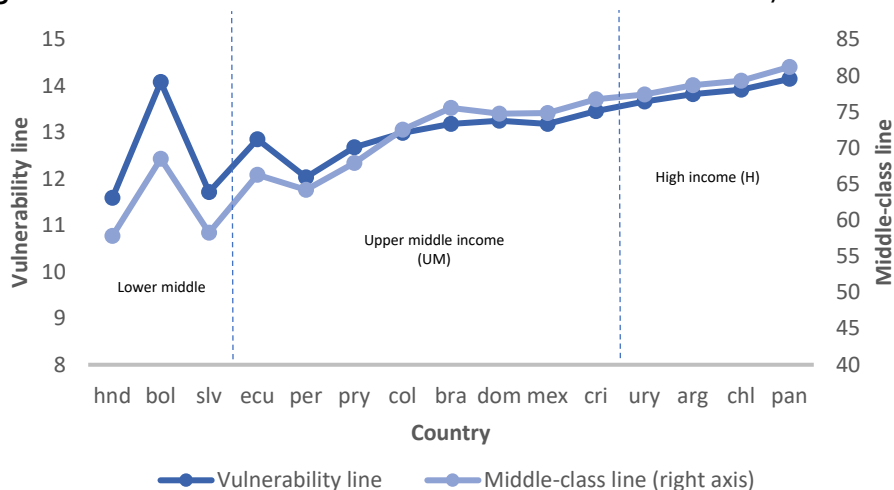
Source: Own estimations based on SEDLAC (2022)
Note: World Bank Analytical Classifications using data for 2017. Results were obtained using 38 synthetic panels built with available surveys within +/- 2 years circa 2017 and 100 bootstrap repetitions.

Robustness check: Increasing the time span

When estimating the International Poverty Lines, Jolliffe et al. (2020) select for each country one survey that was conducted in 2017 or the closest year. The proposed analysis in this paper followed the same principle by choosing a two-year interval around 2017 (i.e., 2015 and 2019) for LAC countries to capture their income and poverty dynamics.[14] However, expanding the time span - including more rounds- might increase the statistical support by embracing changes that are not necessarily assigned to a specific period. To check the robustness of both thresholds, new two-year synthetic panel rounds were constructed for the 2010–2015 interval. Figure 3 shows that when using more than 90 synthetic panels (2010-2019), variations in bounds estimates are not significant: the median lower-middle class line

---

[14] Lopez-Calva & Ortiz-Juarez (2014) exploit the longitudinal data for 3 or 4-year intervals: Chile (2001-2006), Mexico (2002-2005), and Peru (2002-2006).

13

rounds to $14, and the 99th percentile rounds to $80. This suggests that using the two-year interval around 2017 is robust to using a wider period range.

Figure 3: Middle-class lower and upper bound for LAC countries in 2017 PPP

Panel A: Regional vulnerability line   Panel B: Regional middle-class lines
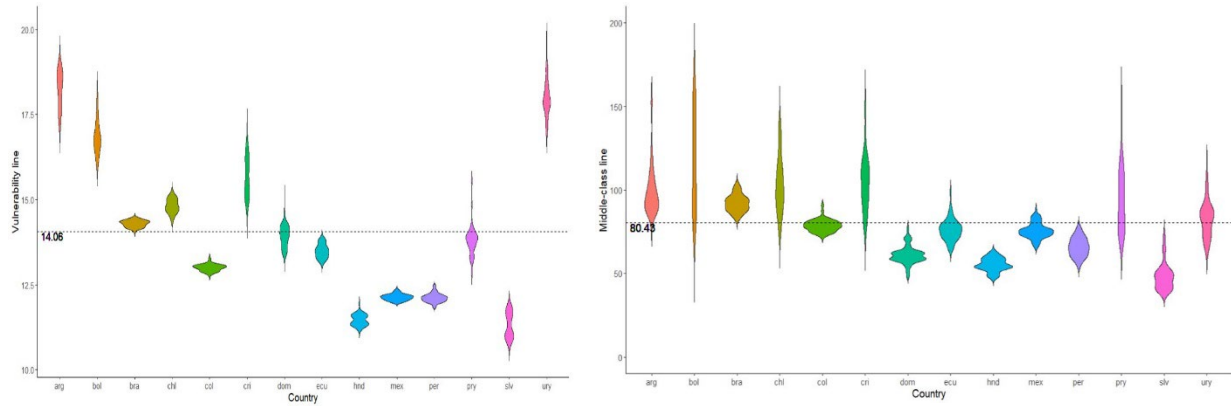
Source: Own estimations based on SEDLAC (2022)
Note: the graphs present confidence intervals for each country based on 25 bootstrap repetitions

Robustness check: Changing the number of bootstrap repetitions

Results might be sensitive to the number of repetitions of the LASSO-PMM method. To tackle this issue, the analysis implemented robustness exercises by decreasing the number of bootstrap repetitions from 100 to 25 to assess whether these lines (i.e., lower and upper bounds) vary significantly or not. The exercise was performed in 15 countries at two timespan intervals (i.e., between 2010 and 2019, and 2015 and 2019). As shown in Figures 1 and 3, the lower and upper middle-class thresholds are quite stable regarding the increase in the repetition number. The lower bound remains around the $14 per person a day (2017 PPP) and the upper bound within the $81 a day (2017 PPP) interval. Figure 4 shows that both vulnerability and middle-class lines estimates are robust to using fewer repetitions (although, as expected, a slightly higher volatility is observed). Each point in the figure corresponds to the line that was estimated using the synthetic panels available for that country and all those to the left.

Figure 4: Cumulative middle-class bounds for LAC countries in 2017 PPP ordered by country GDP
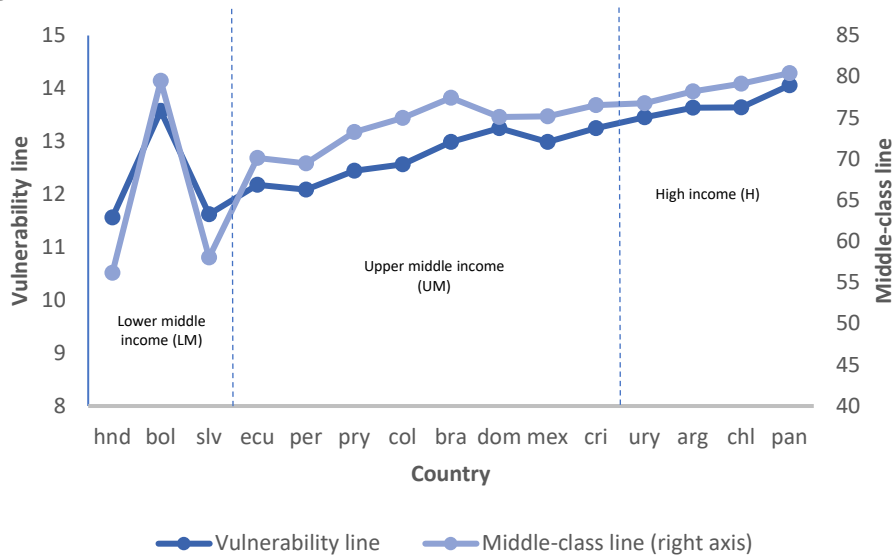
Note: World Bank Analytical Classifications using data for 2017. Results were obtained using 92 synthetic panels built with available surveys from 2010 to 2019 and 25 bootstrap repetitions.

Robustness check: Implementing other methodologies

For completeness, the analysis estimates the middle-class lines using previous methodologies. If the Lopez-Calva and Ortiz-Juarez (2014) methodology is applied only to Chile, Mexico, and Peru -as in the original study- or to Argentina, Colombia, and Costa Rica -as in Ferreira et al. (2013)-, the lower bound would be $12 and $16.1, respectively, and the upper bound $85.8 and $109.3, correspondingly. If this method is expanded to all countries over the same period, the lower bound would end up being $13.6 and the upper bound $89.1. Moreover, replicating the simple method applied, the lower and upper bounds would be $14.7 and $79.3, respectively.

Identifying the drivers of change in the thresholds

While the update of PPPs from 2011 to 2017 contributed to the observed changes in the middle-class thresholds ($ 14 and $ 81) compared to the previous lines ($ 13 and $ 70, respectively), it is not the only factor. One way to initially assess the impact of updating purchasing power parity (PPP) rates is to calculate the ratio (δ) of the relative change in the consumer price index (CPI) between 2011 and 2017 to the relative change in PPP between those same years:

$$\delta = \frac{CPI2017}{CPI2011} \Big/ \frac{PPP2017}{PPP2011}$$

By averaging the values of δ across all 15 countries, it is possible to estimate the impact of the PPP update on the lower and upper thresholds: $\bar{\delta} = 1.13$. On the other hand, the update in the lower threshold from $ 13 in 2011 PPP to $ 14 in 2017 PPP implies a 7.7% increase, and the change in the upper threshold results in a 15.7% increase. In conclusion, the combined effect of all other factors is -5.3% in the lower threshold and an additional 2.7% in the upper threshold.

The methodological approach presented in this paper differs from previous work in three ways: i) the use of synthetic panels instead of actual panels, ii) the definition of the thresholds, and iii) the inclusion of more countries (15 rather than 3 in the original study). To assess the impact of these differences on the real value (aside from the effect of the PPP update) of the new lines, some additional calculations were done. When applying the proposed methodology to the three original countries (Chile, Mexico, and Peru) using 2011 PPPs, the lower and upper bounds would be 3.8 and 11.4% higher, compared to the $ 13 and $ 70 thresholds, respectively (Table 3). This is the combined effect of working with synthetic panels and introducing a new definition of thresholds. On the other hand, when this same analysis is conducted across all 15 available countries, the lower and upper bounds would be 3.8 lower and 3.4% higher, respectively. As a result, it could be concluded that the overall effect of expanding the universe of analysis from 3 to 15 countries is a decrease of 7.4 and 7.2% in the lower and upper thresholds, respectively.

Table 3. Decomposition of changes to middle-class thresholds

Thresholds in 2011 PPP	Original	Proposed approach with 3 countries as in Lopez-Calva & Ortiz-Juarez (2014)	Proposed approach with all 15 countries available
Lower	13	13.5	12.5
Upper	70	78	72.4

Source: Own estimations based on SEDLAC (CEDLAS and the World Bank).

V.   Impacts of 2017 PPPs on poverty, vulnerable, and the middle class

In this section, regional, sub-regional and country-level poverty, vulnerable, and middle-class estimates are presented over time with the 2017 PPPs, in comparison with the 2011 PPPs. This analysis helps understand how incidences and the geographic distribution of the poor, vulnerable, and middle-class populations would change when using 2017 PPPs and higher thresholds. Profiles for these populations are presented over time with the 2017 PPPs, relative to the 2011 PPPs. It is important to assess whether this update of PPPs and thresholds significantly impacts the characteristics of these populations. Marked changes in what the poor and the vulnerable look like affect policy design.

Regional and country-level incidences

Figure 5 illustrates LAC's poverty trends between 2000 and 2020 with the two PPPs sets. Extreme poverty is measured as the proportion of the population living on less than $1.90 or $2.15 a day expressed in 2011 PPP or 2017 PPP. Similarly, graphs in panels B and C show the same incidence of poverty under the $3.2 or $3.65 a day and $5.5 or $6.85 a day expressed in 2011 PPP or 2017 PPP, respectively.

At the regional level, the change from 2011 PPPs to 2017 PPPs induce a relatively small change in extreme poverty and poverty at $3.2 or $3.65 a day in 2011 PPP or 2017 PPP. The 2017 PPP slightly increased historical estimates by less than 0.3 percentage points at $2.15 and 0.9 percentage points at $3.65 in 2017. These growths represent, for instance, 1.9 and 4.8 million additional people in extreme poverty and poverty at $3.65 a day in 2017 PPP. There has been progress in reducing poverty across the region since 2000. Poverty downward trends are similar for both 2011 and 2017 PPPs irrespective of the threshold.

While there are no significant changes in poverty levels at both $2.15 and $3.65 a-day lines (2017 PPP), poverty increases markedly in the LAC region when using the $6.85 line (2017 PPP) relative to the $5.5 line (2011 PPP).[15] In 2017, regional poverty changed by 5 pp, or 27 million more poor people, with the 2017 PPP. The largest changes in millions of poor in 2017 are observed in Brazil (12 million more poor people), and the Andean sub-region (5.5 million more poor). The change in Andean sub-region is mainly driven by Colombia and Peru.

Figure 5: LAC's poverty trends. 2000-2020

Panel A: Poverty Headcount $1.9 (2011 PPP) and Poverty $2.15 (2017 PPP)
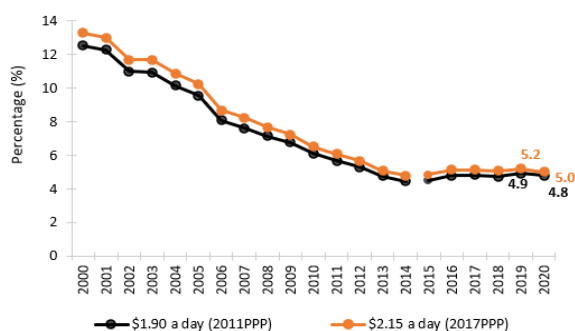
Panel B: Poverty Headcount $3.2 (2011 PPP) and Poverty $3.65 (2017 PPP)

---
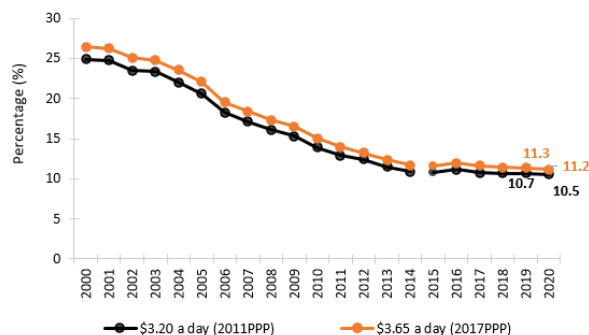
[15] Jolliffe et al (2022) point out that the change in PPPs only accounts for the increase between 5.5 and 6.32, quite far from the final 6.85 line. The relatively high increase in the upper-middle-income line is partially driven by real upward shifts in the national poverty lines of upper-middle-income countries. Part of this can be explained by some of these countries now being high-income countries (for further details in this regard see Jolliffe et al 2022).

Panel C: Poverty Headcount $5.5 (2011 PPP) and Poverty $6.85 (2017 PPP)
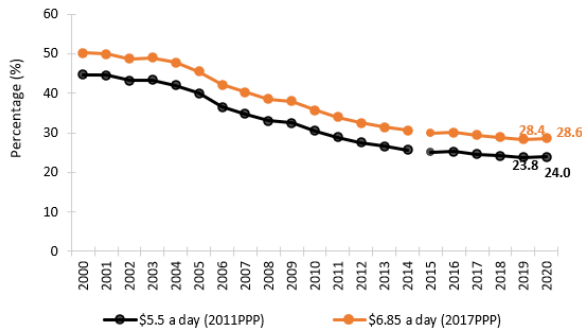
Source: Own estimations based on SEDLAC (CEDLAS and the World Bank).
Note: The LAC aggregate is based on 18 countries in the region for which microdata are available. In cases where data are unavailable, values have been interpolated or extrapolated using WDI data and then pooled to create regional estimates (2014 backward) and microsimulations (from 2015 onwards). Due to important methodological changes in Mexico's official household survey in 2016 that created a break in the poverty series, we have created a break in the LAC-18 aggregate. Version: October 03, 2022.

The significant increase in poverty at $6.85 a day line (2017 PPP) couples with the opposite trend in the region's vulnerable population. The vulnerable population is measured as the share of the population living between $5.5 and $13 a day or $6.85 and $14 a day expressed in 2011 PPP or 2017 PPP, respectively. Overall, vulnerability decreases by 5.4 percentage points with the 2017 PPP (Figure 6), which represents 31 fewer vulnerable. This group reduces substantially in Mexico, the Andean region, and Central America, driving down the regional count of the vulnerable populations by 175 million in 2017. Brazil also experiences a noticeable reduction in its vulnerable population and contributes 29 percent (9 million fewer vulnerable people) of the regional vulnerable population.

Figure 6: LAC's vulnerability and middle-class trends. 2000-2020

Panel A: Vulnerable $5.5-13 (2011PPP) and Vulnerable $6.85-14 (2017PPP)
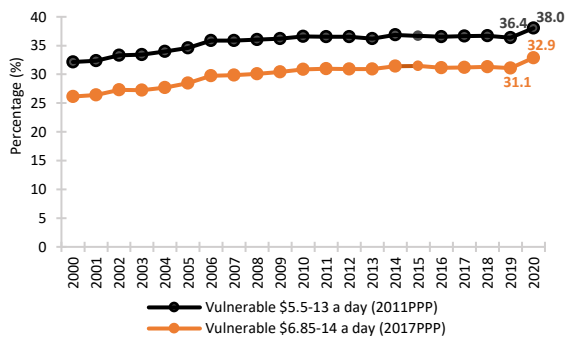
Panel B: Middle-Class $13-$70 (2011 PPP) and Middle- Class $14-$81 (2017 PPP)
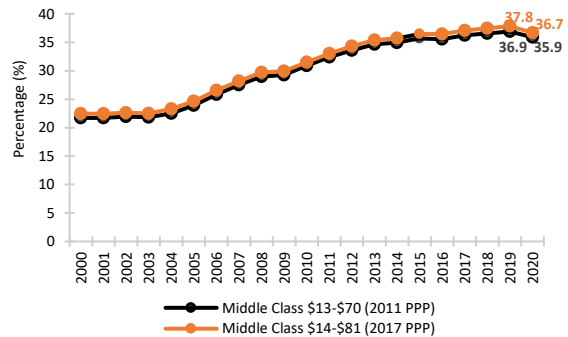
Source: Own estimations based on SEDLAC (CEDLAS and the World Bank).
Note: The LAC aggregate is based on 18 countries in the region for which microdata are available. In cases where data are unavailable, values have been interpolated or extrapolated using WDI data and then pooled to create regional estimates (2014 backward) and microsimulations (from 2015 onwards). Due to important methodological changes in Mexico's official household survey in 2016 that created a break in the poverty series, we have created a break in the LAC-18 aggregate. Version: October 03, 2022.

To understand the sub-regional movements further, Table 4 shows the three countries with the largest absolute changes in vulnerability. The decrease observed in the Andean region is driven by Colombia, where the vulnerable population falls by 6 pp (equivalent to 3 million fewer vulnerable people). About 31 percent of the change in Central America is driven by Guatemala, where vulnerability decreases by 6.8 pp (equivalent to 1 million fewer vulnerable).

Table 4. Countries with the largest absolute changes in vulnerability. Millions of people

|	2015	2016	2017	2018	2019	2020
LAC	*-29*	*-30*	*-31*	*-30*	*-30*	*-30*
Brazil	-8	-9	-9	-9	-9	-7
Mexico*	-8	-8	-8	-9	-9	-9
Andean Region						
Colombia	-3	-3	-3	-3	-3	-3
Peru	-2	-1	-2	-2	-2	-2
Ecuador	-1	-1	-1	-1	-1	-1
Central America						
Guatemala*	-1	-1	-1	-1	-1	-1
Dominican Republic	-1	-1	-1	-1	0	-1
El Salvador	-1	-1	-1	-1	-1	-1

Source: Own estimations based on SEDLAC (CEDLAS and the World Bank).
*For Mex: Projections for 2015, 2017, and 2019. For Guatemala: Projections from 2015 onwards

The middle class has slightly increased in the region and remains the largest socioeconomic group in the LAC region since 2010. This group is measured as the share of the total population between $13 and $70 a day or $14 and $81 a day expressed in 2011 PPP or 2017 PPP, respectively. Between 2010 and 2020, changes in the middle class at the LAC level are relatively small as increases in the three subregions (i.e., Andean, Central America, and Southern Cone) and Mexico are offset by Brazil. In 2017, the middle class in the region grew by less than 1 pp (0.9 pp), or 4.8 million more middle-class people, with the 2017 PPPs. The largest changes in millions of middle-class people in 2017 are observed in the Southern Cone (2.2 million more people) and Brazil (2 million fewer people). The change in the Southern Cone is mainly driven by Argentina.

Regional and country-level profiles

This section focuses on assessing regional profiles of the poor at the upper-middle-income line ($6.85 in 2017 PPP), the vulnerable, and middle-class populations. Even though a significant shift in poverty impacts poverty profiles, these movements are not big enough for inducing changes in policy choices (see Table 5). For instance, the poor are still more concentrated in urban areas, belong to the cohort between 15 and 64 years old, are slightly more educated, work in services, and are less self-employed. Similarly, vulnerability decreases markedly with the new thresholds, but the characteristics of vulnerable people remain the same.[16] The majority of the vulnerable are settled in urban areas, a little bit older by

---

[16] Variations are not statistically significant.

slightly increasing the share of the 15 to 64 y.o. cohort, somewhat more educated, work mainly in services, and a little bit more as salaried workers. Finally, the profiles for the middle-class population were not statistically affected by these changes.

Table 5. Profile of the Poor

		Poverty $5.5 (2011 PPP)						Poverty $6.85 (2017 PPP)						Diff					
		2015	2016	2017	2018	2019	2020	2015	2016	2017	2018	2019	2020	2015	2016	2017	2018	2019	2020
% Population		25.1	25.3	24.6	24.2	23.8	24.0	29.9	30.0	29.5	28.9	28.4	28.6	5	5	5	5	5	5
Area	Urbano	59	59	59	59	60	62	61	61	61	61	63	64	2	2	3	2	2	2
	Rural	41	41	41	41	40	38	39	39	39	39	37	36	-2	-2	-3	-2	-2	-2
Gender	Males	48	48	48	48	48	48	48	48	48	48	48	48	0	0	0	0	0	0
	Females	52	52	52	52	52	52	52	52	52	52	52	52	0	0	0	0	0	0
Age Group	0-14 Years Old	39	39	38	38	37	37	38	38	37	37	36	36	-1	-1	-1	-1	-1	-1
	15-64 Years Old	56	57	57	57	58	59	57	58	58	58	59	59	1	1	1	1	1	1
	65 and older	5	4	5	5	5	5	5	5	5	5	5	5	0	0	0	0	0	0
Education	Years of education	5	5	5	5	5	6	5	5	5	5	5	6	0	0	0	0	0	0
	Less than primary	57	57	57	56	54	48	56	56	56	55	53	48	-1	-1	-1	-1	-1	0
	Primary & less than secondary	32	31	31	31	32	34	31	31	31	31	31	33	0	0	0	0	0	-1
	Secondary	10	11	11	12	13	15	12	12	13	13	14	16	1	1	1	1	1	1
	Tertiary	1	1	1	1	1	2	1	1	1	1	1	2	0	0	0	0	0	0
Sectors	Agriculture	39	39	39	38	37	38	35	35	35	35	34	36	-3	-3	-3	-3	-3	-3
	Industry	20	19	19	19	19	19	20	20	20	20	20	19	1	1	1	1	1	0
	Services	42	42	42	42	44	43	45	45	45	45	46	45	3	3	3	3	2	2
Type of worker	Employers	5	5	5	5	5	6	5	4	4	5	5	5	0	0	0	0	0	0
	Salaried workers	42	40	39	39	39	37	44	43	41	41	41	39	3	2	3	2	2	2
	Self-employed	30	30	30	30	29	29	29	29	29	29	29	28	-1	-1	-1	-1	0	-1
	Not salaried	11	11	11	11	11	12	10	10	10	10	10	11	-1	-1	-1	-1	-1	-1
	Unemployed	12	14	16	15	16	17	11	14	15	15	16	17	0	0	0	0	0	0
Household characteristics and access to services	Average hh size	5	5	5	5	5	5	5	5	5	5	5	5	0	0	0	0	0	0
	Access to electricity (%)	97	98	98	98	98	97	98	98	98	98	98	98	0	0	0	0	0	0
	Access to improved water (%)	87	89	89	90	90	90	88	90	90	91	91	90	1	1	1	1	1	1

Source: SEDLAC (World Bank and CEDLAS).

Note: Note: Since the numbers presented here are based on SEDLAC, a regional data harmonization effort that increases cross-country comparability, they may differ from official statistics reported by governments and national statistical offices. The LAC aggregate is based on 18 countries in the region for which microdata are available. In cases where data are unavailable for a given country in a given year, values have been interpolated using WDI data or using microsimulations to calculate regional measures. Type of employment and sector limited to working individuals ages 15–64.

Table 6. Profile of the Vulnerable

		Vulnerable $5.5-$13 (2011 PPP)						Vulnerable $6.85-$14 (2017 PPP)						Diff					
		2015	2016	2017	2018	2019	2020	2015	2016	2017	2018	2019	2020	2015	2016	2017	2018	2019	2020
% Population		36.7	36.5	36.5	36.6	36.1	37.7	31.4	31.1	31.1	31.2	30.8	32.6	-5.2	-5.4	-5.4	-5.4	-5.3	-5.1
Area	Urbano	80	80	80	80	80	80	82	81	81	81	80	81	1	1	1	1	1	1
	Rural	20	20	20	20	20	20	18	19	19	19	20	19	-1	-1	-1	-1	-1	-1
Gender	Males	48	48	49	48	48	48	48	48	48	48	48	48	0	0	0	0	0	0
	Females	52	52	51	52	52	52	52	52	52	52	52	52	0	0	0	0	0	0
Age Group	0-14 Years Old	28	27	27	27	26	26	27	26	26	26	26	25	-1	-1	-1	-1	-1	-1
	15-64 Years Old	66	66	66	66	66	66	67	67	67	67	67	67	1	1	1	1	0	1
	65 and older	7	7	7	7	8	7	7	7	7	7	8	8	0	0	0	0	0	0
Education	Years of education	7	7	7	7	7	7	7	7	7	7	7	7	0	0	0	0	0	0
	Less than primary	42	42	41	40	40	39	41	41	40	40	39	39	-1	-1	-1	-1	-1	0
	Primary & less than secondary	32	32	32	32	32	30	32	32	32	32	31	29	0	0	0	0	0	-1
	Secondary	22	23	23	24	25	26	23	24	24	24	25	27	1	1	1	1	1	1
	Tertiary	3	3	4	4	4	5	4	4	4	4	4	5	0	0	0	0	0	0
Sectors	Agriculture	13	13	13	14	14	15	12	12	13	13	13	14	-1	-1	-1	-1	-1	-1
	Industry	25	25	24	24	24	24	25	25	24	24	24	24	0	0	0	0	0	0
	Services	62	62	62	62	62	61	62	63	63	63	63	62	1	1	1	1	1	1
Type of worker	Employers	3	3	3	3	4	3	3	3	3	3	3	3	0	0	0	0	0	0
	Salaried workers	63	61	60	60	59	57	64	62	62	61	60	58	1	1	1	1	1	1
	Self-employed	22	23	23	23	24	23	22	22	23	23	23	22	0	0	0	0	0	0
	Not salaried	4	4	4	4	4	5	4	4	4	4	4	4	0	0	0	0	0	0
	Unemployed	8	9	9	9	10	13	7	9	9	9	9	13	0	0	0	0	0	0
Household characteristics and access to services	Average hh size	5	4	4	4	4	4	4	4	4	4	4	4	0	0	0	0	0	0
	Access to electricity (%)	99	99	99	99	99	99	99	99	99	99	99	99	0	0	0	0	0	0
	Access to improved water (%)	96	97	97	97	97	97	96	97	97	97	97	97	0	0	0	0	0	0

Source: SEDLAC (World Bank and CEDLAS).

Note: Note: Since the numbers presented here are based on SEDLAC, a regional data harmonization effort that increases cross-country comparability, they may differ from official statistics reported by governments and national statistical offices. The LAC aggregate is based on 18 countries in the region for which microdata are available. In cases where data are unavailable for a given country in a given year, values have been interpolated using WDI data or using microsimulations to calculate regional measures. Type of employment and sector limited to working individuals ages 15–64.

Table 7. Profile of the Middle Class

		Middle Class $13-$70 (2011 PPP)						Middle Class $14-$81 (2017 PPP)						Diff					
		2015	2016	2017	2018	2019	2020	2015	2016	2017	2018	2019	2020	2015	2016	2017	2018	2019	2020
% Population		35.8	35.7	36.3	36.6	37.6	36.0	36.5	36.6	37.2	37.6	38.6	36.8	0.8	0.9	0.9	0.9	1.0	0.8
Area	Urbano	90	90	90	90	90	89	90	90	90	90	90	89	0	0	0	0	0	0
	Rural	10	10	10	10	10	11	10	10	10	10	10	11	0	0	0	0	0	0
Gender	Males	50	49	49	49	49	49	49	49	49	49	49	49	0	0	0	0	0	0
	Females	50	51	51	51	51	51	50	51	51	51	51	51	0	0	0	0	0	0
Age Group	0-14 Years Old	15	15	15	14	14	14	15	15	15	15	14	14	0	0	0	0	0	0
	15-64 Years Old	73	72	72	72	72	72	73	72	72	72	72	72	0	0	0	0	0	0
	65 and older	12	13	13	13	14	14	12	12	13	13	14	14	0	0	0	0	0	0
Education	Years of education	9	9	9	9	9	9	9	9	9	9	9	10	0	0	0	0	0	0
	Less than primary	28	27	27	27	26	27	28	27	27	27	26	27	0	0	0	0	0	-1
	Primary & less than secondary	23	23	22	22	22	20	23	23	23	22	22	21	1	1	1	1	1	1
	Secondary	33	33	33	33	34	34	32	33	33	33	33	33	0	0	0	0	0	0
	Tertiary	16	17	17	18	18	19	16	17	17	18	18	19	0	0	0	0	0	0
Sectors	Agriculture	5	5	6	6	6	7	5	5	5	6	6	7	0	0	0	0	0	0
	Industry	22	21	21	20	20	20	22	21	21	20	20	21	0	0	0	0	0	0
	Services	73	74	74	74	74	73	73	74	74	74	74	73	0	0	0	0	0	0
Type of worker	Employers	5	5	5	5	5	5	5	5	5	5	5	5	0	0	0	0	0	0
	Salaried workers	70	69	69	69	68	67	70	69	69	69	68	67	0	0	0	0	0	0
	Self-employed	19	19	20	20	20	20	19	19	20	20	20	20	0	0	0	0	0	0
	Not salaried	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0
	Unemployed	4	5	5	4	5	6	4	4	5	4	5	6	0	0	0	0	0	0
Household characteristics and access to services	Average hh size	3	3	3	3	3	3	4	4	3	3	3	3	0	0	0	0	0	0
	Access to electricity (%)	99	100	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0
	Access to improved water (%)	99	99	99	99	99	99	98	99	99	99	99	99	0	0	0	0	0	0

Source: SEDLAC (World Bank and CEDLAS).

Note: Note: Since the numbers presented here are based on SEDLAC, a regional data harmonization effort that increases cross-country comparability, they may differ from official statistics reported by governments and national statistical offices. The LAC aggregate is based on 18 countries in the region for which microdata are available. In cases where data are unavailable for a given country in a given year, values have been interpolated using WDI data or using microsimulations to calculate regional measures. Type of employment and sector limited to working individuals ages 15–64.

VI.   Final remarks

This study analyzes the impacts of adjusting the LAC lower and upper-middle-class regional thresholds when new and better rounds of PPPs are estimated. First, the paper proposes a two-step methodology to update these thresholds. The first step consists in building two-year synthetic panels between 2015 and 2019 for 15 countries in the LAC region. It is based on the paper by Lucchetti et al. (2020) and introduces several improvements in the methodology: matching is done using constrained donation subsets based on gender, birth year, and educational attainment; introduces weights harmonization procedure to preserve the original distribution in terms of relevant variables for the analysis; neighbor selection is based on random selection within the subset of closest neighbors, considering the sampling design of the donor survey; it implements more bootstrap repetitions to achieve more accuracy in estimates, etc. The second step defines the *lower bound* as the median per capita household income over households who were not poor in the initial year and became poor in the final year and the *upper bound* as the maximum per capita income (99th percentile) over those who have changed their status. This yields a vulnerability or lower-bound line of $14 in 2017 PPP per person per day and a middle-class or upper line of $81 in 2017 PPP. These thresholds are robust to expanding the set of countries, widening the time interval, increasing the number of repetitions, and implementing previous methodologies.

This study analyzes the impact on the LAC regional poor, vulnerable, and middle-class groups estimate using the international poverty lines (IPL), the regional vulnerability and middle-class lines, and the 2017 PPPs. While the changes at the upper-middle-income line ($6.85 in 2017 PPP) are significantly larger, the 2017 PPPs have small implications for extreme poverty and poverty at the lower-middle-income line ($2.15 and $3.65 lines). Extreme poverty and poverty at $3.65 would increase marginally by 0.3 and 0.9 percentage points (pp.), respectively in 2017, with the 2017 PPPs. The regional count of the extreme poor increased by 1.9 million, largely driven by 1.3 million and 300,000 more poor people in Brazil and the Andean region, while poverty increases in all other subregions except the Southern Cone. Considering the upper-middle income country line, the incidence of poverty in the region rose around 5 percentage points in 2017 with the 2017 PPP. This represents 27 million more poor people in LAC driven by 12 million and 5.5 million in Brazil and the Andean region, respectively.

A significant increase in the poor population is mirrored by a decrease in the size of the vulnerable group by around 5.4 pp with the revised lower-middle-class bound of $14 in 2017 PPP. The regional vulnerability trend is due to declines in the Andean region and Central America offset by increases in countries in the Southern Cone, including Brazil, since 2012. With the update of the upper-middle-class threshold of $81 in 2017 PPP, the size of the middle class slightly increases regionally by less than 1pp and remains the largest socioeconomic group in the LAC region since 2010. Between 2010 and 2020, changes in the middle class at the LAC level are relatively small as increases in the three subregions (i.e., Andean, Central America, and Southern Cone) and Mexico are offset by Brazil. This change in lower and upper middle-class limits provides a similar positive trend for this group over the last two decades. More importantly, these updates in thresholds do not significantly impact the profiles of the poor, the vulnerable, or the middle-class populations, which yield similar policy choices.

Using the approach based on the risk of falling into poverty, as proposed by Lopez-Calva & Ortiz-Juarez (2014), is conceptually more accurate in defining the lower bound of the middle class for country-specific analysis. Interestingly, it has been found that despite using an alternative approach, this paper's results align closely with their findings, which is a valuable contribution to the discussion on middle-class thresholds. Moreover, while the risk-based approach is highly relevant and useful for defining the lower end of the middle class, it does not address the upper end of the spectrum. The proposed methodology in this paper simultaneously tackles the lower and upper bounds of the middle class, using a consistent methodological approach. Doing so provides a comprehensive and nuanced understanding of the middle class in Latin America, which can be of great value to policy makers and researchers alike.

Despite offering valuable insights into the socioeconomic structure of the region, this methodology has certain limitations that should be acknowledged. First, the synthetic panel construction using labor market surveys is subject to potential biases and measurement errors inherent in self-reported income data. These biases may impact the accuracy of the estimated bounds for the middle class. Additionally, the synthetic panel data may not capture the full extent of income mobility based on cross-sectional surveys rather than longitudinal data. Second, this approach does not explicitly account for factors such as social protection systems, access to quality education and health care, or other non-income dimensions contributing to an individual's vulnerability or resilience to falling into poverty. As a result, the proposed methodology may not fully capture the complex dynamics that define the middle class in Latin America. It may overlook essential aspects related to social inclusion, economic security, and quality of life. Lastly, this method may need to comprehensively understand the heterogeneity within the middle class. This group is diverse, encompassing households with varying levels of vulnerability. This approach may not adequately capture these nuances, which are essential for designing effective policies to foster inclusive growth.

While the present approach offers a sound perspective on defining the middle class in Latin America, it is important to recognize its limitations and consider complementary methodologies to better understand the complex dynamics of this socioeconomic group. By acknowledging these limitations, this paper aims to contribute to a richer and more informed debate on the region's identification and characterization of the middle class.

References

Balcazar, C.; Dang, H.; Malasquez, E.; Olivieri, S.; Pico, J. (2018). Welfare Dynamics in Colombia: Results from Synthetic Panels. Policy Research Working Paper; No. 8441. World Bank, Washington, DC.

Birdsall, N., F. Ferreira, L.F. Lopez-Calva & J. Rigolini (2011). ―The Middle Class in Developing Countries: Concept, Measurement, and Recent Trends, Mimeo. World Bank, Washington, D.C.

Bourguignon, F. (2015) Appraising Income Inequality Databases in Latin America. *Journal of Economic Inequality* 13(4): 557–78.

Bourguignon, F. and Moreno, H. (2015). "On the construction of synthetic panels". MIMEO. NEUDC. October 2015.

Dang, H, Lanjouw, P., Luoto, J., and McKenzie. D. (2014). "Using Repeated Cross-Sections to Explore Movements in and out of Poverty". Journal of Development Economics, 107: 112-128.

Dang, H. and Lanjouw. P. (2013). "Measuring poverty dynamics with synthetic panels based on cross-sections." World Bank Policy Research Working Paper 6540

Dang, H. and Lanjouw. P. (2016). "Measuring Poverty Dynamics with Synthetic Panels Based on Repeated Cross-Sections." Available at: http://lacer.lacea.org/handle/123456789/61406

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.

Ferreira, F.; Messina, J.; Rigolini, J.; López-Calva, L.; Lugo, M.; Vakis, R. (2013) Economic Mobility and the Rise of the Latin American Middle Class. World Bank Latin American and Caribbean Studies; Washington, DC

Jolliffe, D., and Prydz, E. (2016). "Estimating International Poverty Lines from Comparable National Thresholds." Journal of Economic Inequality 14: 185–98. https://doi.org/10.1007/s10888-016-9327-5

Jolliffe, D., Mahler, D., Lakner, C., Atamanov, A., and Tetteh-Baah, S. (2022) Assessing the Impact of the 2017 PPPs on the International Poverty Line and Global Poverty. Policy Research Working Paper 9941, World Bank, Washington, D.C.

López-Calva, L., Ortiz-Juarez, E. (2014). "A vulnerability approach to the definition of the middle class". Journal of Economic Inequality 12: 23–47. https://doi.org/10.1007/s10888-012-9240-5.

Lucchetti, L., Corral, P., Ham, A., & Garriga, S. (2020). Lassoing Welfare Dynamics with Cross-Sectional Data. Policy Research Working Paper 8545. World Bank. Washington, D.C.

Ñopo, H. (2004). Matching as a tool to decompose wage gaps, IZA. Discussion Papers, No. 981, Institute for the Study of Labor (IZA), Bonn.

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.

Renssen R.H. (1998). "Use of statistical matching techniques in calibration estimation". Survey Methodology 24, 171-183.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

World Bank (2021) The Gradual Rise and Rapid Decline of the Middle Class in Latin America and the Caribbean. World Bank, Washington, DC.