

# Considering Labor Informality in Forecasting Poverty and Inequality

A Microsimulation Model for Latin American  
and Caribbean Countries

*Kelly Montoya*

*Sergio Olivieri*

*Cicero Braga*



**WORLD BANK GROUP**

Poverty and Equity Global Practice

June 2023

## Abstract

Economists have long been interested in measuring the poverty and distributional impacts of macroeconomic projections and shocks. In this sense, microsimulation models have been widely used to estimate the distributional effects since they allow accounting for several transmission channels through which macroeconomic forecasts could impact individuals and households. This paper innovates previous microsimulation methodology by introducing more flexibility in labor earnings, considering intra-sectoral variation according to the formality status, and assessing

its effect on forecasting country-level poverty, inequality, and other distributive indicators. The results indicate that the proposed methodology accurately estimates the intensity of poverty in the most immediate years indistinctively of how labor income is simulated. However, allowing for more intra-sectoral variation in labor income leads to more accurate projections in poverty and across the income distribution, with gains in performance in the middle term, especially in atypical years such as 2020.

---

This paper is a product of the Poverty and Equity Global Practice. 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/prwp>. The authors may be contacted at [solivieri@worldbank.org](mailto: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.*

# Considering Labor Informality in Forecasting Poverty and Inequality: A Microsimulation Model for Latin American and Caribbean Countries<sup>1</sup>

Montoya, Kelly, Olivieri, Sergio and Braga, Cicero<sup>2</sup>

JEL: D39, E27, O15

*Keywords: micro-simulation, ex-ante, poverty, inequality, forecasting, Latin America and the Caribbean*

---

<sup>1</sup> We are grateful for the multiple interactions and discussions with economists and consultants from the Poverty and Equity Practice in the Latin America and Caribbean region (ELCPV): Carolina Mejia-Mantilla, Gabriel Lara-Ibarra, Ricardo Vale, Gabriela Inchauste, Monica Robayo, Jacobus Hoost de Hoop, Agustin Arakaki, Javier Romero, Gustavo Canavire, Lourdes Rodriguez, Trinidad Saavedra, Alejandro de la Fuente, Hernan Winkler, Maria Davalos, Ana Rivadeneira, and Erika Schutt, which enriched the paper. The paper benefitted from feedback from reviewers Hugo Ñopo, and Samuel Freije-Rodriguez, additional comments from Christoph Lakner, seminar participants at the World Bank workshops, and the guidance of Ximena del Carpio and Carlos Rodriguez-Castelan.

<sup>2</sup> Correspondence author: [solivieri@worldbank.org](mailto:solivieri@worldbank.org)

## Introduction

Economists have long been interested in measuring the poverty and distributional impacts of macroeconomic projections based on structural reforms, macroeconomic shocks, and other events. A standard solution is to extrapolate the welfare impact of these projections from the historical responses of income (consumption) poverty to changes in output by estimating an elasticity of poverty to output or gross domestic product (GDP). Although this approach is easy and rapid to implement, it is limited in its predictive capability since it cannot estimate distributional impacts (i.e., poverty gap, inequality, vulnerability, etc.).

To estimate distributional impacts, microsimulation models allow accounting for several transmission channels through which macroeconomic projections could impact individuals and households. Thus, it helps evaluate the consequences of a change in the economic environment induced by a macroeconomic scenario on the welfare of each individual or household and identifies those likely to be losers and winners. The more sophisticated microsimulation models are based on computable general equilibrium (CGE) or general equilibrium macroeconomic models that demand substantial information (for constructing social accounting matrices or time series of macroeconomic data) to create the "linkage aggregate variables" (LAVs) that are fed into the microsimulation model (Bourguignon et al., 2008). At the same time, most of these models do not allow for changes in some key features of the population, such as gender or age composition, except for the Maquette for MDG [Millennium Development Goal] Simulations (MAMS).<sup>3</sup> Further, the relationship between the CGE model and the microsimulation model can be sequential (top-down approach), in which case the outputs of the CGE model are used in the microsimulation as inputs (Bourguignon et al. 2003; Heraoult, 2010); or iterative (top-down/bottom-up approach) in which case the results of the microsimulation model are reincorporated in the CGE model as inputs until an equilibrium between micro and macroeconomic estimations is achieved (Savard, 2003; Colombo, 2010). The main advantages of the iterative models relate to the improved accuracy of the counterfactual and consistency of the analysis. However, the information demands of these models make them difficult to apply in most developing countries, thus calling for an approach that is workable with the available data and macroeconomic projections.

Following the top-down approach, there are multiple types of microsimulation models, from those which ignore behavioral 'agents' responses to changes in the economic environment (arithmetical or accounting approaches) as in Buddelmeyer et al. (2008) and Ferreira and Horridge (2006), to those which include a detailed representation of behavioral responses of individuals or households in aspects such as occupation (Olivieri, S. et al., 2014) or savings behavior (Van Ruijven et al., 2015). There is also a differentiation between static and dynamic micro-simulation models. The former does not consider changes in the baseline sociodemographic characteristics of individuals and households, such as the level of education, the household composition, demographic change, etc.; the latter introduces changes in individuals' and households' sociodemographic characteristics and behavior over time derived from

---

<sup>3</sup> Ferreira et al. 2008.

changes in the macroeconomic environment. Examples of these changes are decisions on training, child conception, etc. (Bourguignon and Spadaro, 2006). The microsimulation model studied in this document belongs to those static with behavioral 'agents' responses.

Among the behavioral models, Olivieri et al. (2014) present a microsimulation model that evaluates the distributional impacts of a macroeconomic shock with low data and computational requirements. It allows accounting for labor and non-labor income mechanisms and captures impacts at the micro level for the entire income distribution. The model focuses on labor market adjustments in employment and earnings and changes in non-labor income and prices (with a view to the variation in food and non-food prices). However, it does not allow for capturing labor informality, an important feature for Latin American countries. Moreover, Olivieri (2020) adapts the previous version of the microsimulation model to assess the effects of the triple crisis in Ecuador. This version incorporates labor informality only in estimating the labor market structure in the three main sectors of economic activity. Yet, its predictive capacity is limited at the labor income level.

In the absence of CGE models, an alternative way to feed microsimulation models under the top-down approach is using macroeconomic projections of output that are almost always available. However, since employment and labor income estimations are usually unavailable, the output growth estimates must be translated into employment and labor income changes at the sector level. These estimates are typically made using sectoral output-employment and productivity-labor income elasticities based on aggregate output and labor market past data, which are then applied to the output growth projections to generate changes in employment and labor income by sector (Braga, C. et al., 2023). The predicted sectoral employment and income, along with the macroeconomic output projections, are the final inputs for the microsimulation model. In this paper, these essential inputs come from actual data to isolate biases from the model and the macroeconomic inputs.

This paper considers the model proposed by Olivieri (2020) and assesses the gains in the predictive capacity of microsimulation models when different assumptions are applied in the way family income components are estimated between 2017 and 2020. This work isolates the effect of changes in the microsimulation model's assumptions by considering the use of actual macroeconomic and labor market input data for the period of analysis. Thus, the study first contributes to extending previous microsimulation methodologies by introducing informality to estimate how labor income moves within each sector of economic activity. Second, it tests the performance of three versions of the model (i.e., the *old model*, the *new model with rescaling*, and the *new model without rescaling*) in fitting the actual income distribution in four consecutive years. The results indicate that overall, the methodology proposed successfully identifies the poor and estimates the intensity of poverty in the most immediate years, indistinctively of how labor income is simulated. However, allowing for more intra-sectoral variation in labor income results in more accurate projections of poverty and changes along the income distribution, with gains in its performance in the middle term, especially in atypical years such as 2020.

The rest of the paper is organized as follows. Section I lays out the methodological approach, which differs from traditional techniques (i.e., the elasticity of poverty to output or GDP) and micro-simulation

methods used in the past. Section II introduces macroeconomic and microeconomic input data used in the analysis. Section III assesses the goodness of fit of three model variations to actual distributions and the effects on poverty, inequality, and growth incidence curves. Section IV presents the final remarks.

## I. Methodological approach

The estimates and analysis presented in this paper use an improved micro-simulation model to predict the welfare and distributional impacts of growth. The micro-simulation model superimposes macroeconomic projections on behavioral models built on the last available household survey for each Latin American and Caribbean country. The model is loosely based on previous approaches to micro-simulation described in Bourguignon et al. (2008) and Ferreira et al. (2008). The main difference here is the omission of the computable general equilibrium (CGE) component, which is challenging to employ in most developing countries.<sup>4</sup> Instead of a CGE, the approach described in this paper links the behavioral microsimulation model to aggregate macroeconomic data for LAC countries. However, since employment and labor income estimations are usually not available, the output growth estimates must be translated into employment and labor income changes at the sector level. These estimates are typically made using sectoral output-employment and productivity-labor income elasticities based on aggregate output and labor market past data, which are then applied to the output growth projections to generate changes in employment and labor income by sector (Braga, C. et al., 2023).<sup>5</sup> The predicted sectoral employment and income, along with the macroeconomic output projections, are the final inputs for the microsimulation model. This approach has been extended to explicitly consider informality within economic sectors, which comprises a major problem in the LAC region.

This micro-simulation model accounts for multiple transmission mechanisms that affect family labor and non-labor income and captures impacts at the micro level across the income distribution. In particular, the model can consider significant changes in population over time; labor market adjustments in employment and earnings, or a combination of both; and changes in non-labor incomes, including international remittances, capital, pensions, and public transfers.

### The micro-simulation model setup<sup>6</sup>

The micro-simulation is divided into three steps: the baseline, the simulation, and the assessment. It is based on Olivieri et al. (2014) and Olivieri, S. (2020), albeit with a major difference in accounting for labor market informality when projecting labor income.

---

<sup>4</sup> For further details on implementing an integrated approach, reach out to the CGE-GIDD – MTI team.

<sup>5</sup> This approach was conceptualized, refined, and tested in a diverse mix of countries during the financial crisis (such as Bangladesh, the Philippines, Mexico, Poland, and Mongolia) as well as after the crisis (such as Costa Rica, Panama, Uruguay, Serbia, Armenia, Belarus, Kyrgyz Republic, Moldova, Poland, Romania, and Ukraine). More recently, it was adjusted for the last twin crises in Iraq.

<sup>6</sup> This section is based on Olivieri, S. et al. (2014).

## Baseline

The first step is the process by which individual and household-level information is used to estimate a set of parameters and unobserved characteristics for various household income generation model equations. The model behind the micro-simulation is the household income generation model developed by Bourguignon and Ferreira (2005). This model allows accounting for multiple transmission channels for both family labor and non-labor income, as well as working at the individual/household level. The first component of the model is an identity that defines the per capita income in a household  $h$  as the ratio between the total household income and the total number of members ( $n_h$ ) in that household:

$$y_h = \frac{1}{n_h} \left[ \sum_{i=1}^{n_h} \sum_{L=1}^{\Lambda} \sum_{j=0}^J I_{hi}^{Lj} y_{hi}^{Lj} + y_{0h} \right] \quad (1)$$

where

- $i$  = household member
- $L$  = level of education
- $\Lambda$  = maximum level of education
- $j$  = labor status
- $J$  = employment sector
- $I_{hi}^{Lj}$  = indicator function of labor status  $j$  of individual  $i$  with a level of education  $L$
- $y_{hi}^{Lj}$  = earnings of individual  $i$  with a level of education  $L$  in employment sector  $j$
- $y_{0h}$  = total non-labor income received by household  $h$

The total household income — the expression in brackets in equation (1) — results from adding two main sources of family income: *labor* and *non-labor* income. At the same time, the total family labor income is the aggregation of earnings in different employment sectors across members.<sup>7</sup> So, it is possible to see not only whether an individual does (or does not) participate in the sector, but also whether that individual receives (or does not receive) wages for that job.

The labor participation model relies on the utility maximization approach developed by McFadden.<sup>8</sup> Assume that the utility ( $U_{hi}^{Lj}$ ) for individual  $i$  of household  $h$ , associated with labor status  $j=0, \dots, J$ , and level of education  $L$ , can be expressed as a linear function of observed individual and household characteristics ( $Z_{hi}^L$ ) and unobserved utility determinants of the occupational status ( $v_i^{Lj}$ ). Furthermore, assume individual  $i$  chooses sector  $j$  (the indicator function  $I_{hi}^{Lj} = 1$ ) if employment sector  $j$  provides the highest level of utility:<sup>9</sup>

$$U_{hi}^{Lj} = Z_{hi}^L \Psi^{Lj} + v_i^{Lj} \quad \text{with } j = 0, \dots, J \text{ and } L = \text{education level} \quad (2)$$

<sup>7</sup> Note that although it is possible to estimate specific models for salaried and non-salaried workers based on the microdata from the household survey, it was not possible in this case to use these models because this information is not generally available from the macro side. Macro-economic projections are calculated mainly for aggregate economic sectors, such as agriculture, industry and services, instead of wages or self-employed, formal and informal sectors.

<sup>8</sup> McFadden (1974).

<sup>9</sup> Bourguignon and Ferreira (2005) say that this interpretation is not fully justified because occupational choices may actually be constrained by the demand side of the market, as in the case of selective rationing, rather than individual preferences.

$$I_{hi}^{Lj} = 1 \text{ if } U_{hi}^{Lj} \geq U_{hi}^{Ll} \quad \text{for all } l = 0, \dots, J, \forall l \neq j \quad (3)$$

Each individual must choose from the following alternatives: being inactive, being unemployed, or being active in an employment sector (i.e., agriculture, formal or informal; industry, formal or informal; and services, formal or informal). The parameters of the occupational decision model can be obtained using a multinomial logit model under the assumption that the unobservables ( $v_i^{Lj}$ ) are identically and independently distributed across choices and individuals, and that they have Type I extreme value distribution (double exponential) with density (pdf) and cumulative functions (cdf) given by:

$$f(v_i^{Lj}) = \exp[-\exp(-v_i^{Lj})] \exp(-v_i^{Lj}) \quad (4)$$

$$F(v_i^{Lj}) = \exp[-\exp(-v_i^{Lj})] \quad (5)$$

The estimation is conducted on all individuals of working age (i.e., between 15 and 64 years old), separating for low and high skill levels. The labor force and employment decisions within the household are modeled only by the inclusion of the household head binary variable and its interactions with gender and marital status. The set of explanatory variables includes not only an individual's sociodemographic characteristics (i.e., age, gender, maximum education level, head of the household or not, education enrollment) but also the household's characteristics (i.e., the presence of public workers, dependency ratio, and geographic area – urban/rural, and the region –). So, the parameters can be estimated, as can the probability of being in each state at the individual level, considering zero as the reference category (inactivity):

$$P_{hi}^{Lj} = \frac{\exp[Z_{hi}^{Lj}(\Psi^{Lj} - \Psi^{L0})]}{1 + \sum_{j=1}^J \exp[Z_{hi}^{Lj}(\Psi^{Lj} - \Psi^{L0})]} \quad \text{for } j = 1, \dots, J \quad (6)$$

$$P_{hi}^{L0} = 1 - \sum_{j=1}^J P_{hi}^{Lj} \quad (7)$$

To estimate the individual utility level of being in each labor state, values for the residual terms were drawn randomly in a way that is consistent with observed occupational choices. Train and Wilson (2008) define the distribution functions of the extreme value errors conditional on the chosen alternative. Assume alternative zero is chosen ( $j = 0$ ) and denotes  $Z_{hi}^{Lj} \hat{\Psi}^{Lj} = V_{hi}^{Lj}$  for  $j = 0 \dots, J$ .<sup>10</sup> Defines  $\hat{V}_{hi}^{0j} = V_{hi}^0 - V_{hi}^j$  and  $D_{hi}^0 = \sum_{j=0}^J \exp(-\hat{V}_{hi}^{0j})$  where  $P_{hi}^0 = 1/D_{hi}^0$  is the logit choice probability. Then, the cdf for the alternative chosen  $v_{hi}^0$  is:

$$F(v_{hi}^0 | \text{alternative 0 is chosen}) = \exp(-D_{hi}^0 \exp(v_{hi}^0)) \quad (8)$$

Calculating the inverse of this distribution:

$$\hat{v}_{hi}^0 = \ln(D_{hi}^0) - \ln(-\ln(\mu)) \quad (9)$$

<sup>10</sup> For simplicity, the L superscript, which refers to skill level of the individual, was temporarily removed.



where  $\mu$  is a draw from a uniform distribution between 0 and 1. Error terms for other alternatives ( $v_{hi}^j$  with  $j \neq 0$ ) must be calculated conditioning on the error terms of the alternative chosen ( $\hat{v}_{hi}^0$ ). The distribution for these errors is:

$$F(v_{hi}^j | \text{alternative } 0 \text{ is chosen}, j \geq 1) = \frac{\exp(-\exp(-v_{hi}^j))}{\exp(-\exp(-(\hat{V}_{hi}^{0j} + \hat{v}_{hi}^0)))} \text{ for } v_{hi}^j < (\hat{V}_{hi}^{0j} + v_{hi}^0) \quad (10)$$

The inverse of this distribution is:

$$\hat{v}_{hi}^j = -\ln(-\ln(m(\hat{v}_{hi}^0)\mu)) \quad (11)$$

where  $m(\hat{v}_{hi}^0) = \exp(-\exp(-(\hat{V}_{hi}^{0j} + \hat{v}_{hi}^0)))$

where  $\mu$  is a draw from a uniform distribution between 0 and 1. The obtained residual terms are fixed for each individual and then used to calculate the behavioral responses given the observed characteristics. Repeating this same method when an alternative other than zero is chosen and using expressions (7) to (11), individual utility levels for each alternative can be calculated as:

$$\hat{U}_{hi}^{Lj} = Z_{hi}^{Lj} \hat{\Psi}_{hi}^{Lj} + \hat{v}_{hi}^{Lj} \text{ for } j \geq 0 \quad (12)$$

The observed heterogeneity in earnings in each employment sector  $j$  can be modeled by a log-linear function of observed individual and household characteristics ( $X_{hi}^L$ ) (e.g., age, gender, type of employment, informality, geographic area, among others) and unobserved factors ( $\mu_{hi}^{Lj}$ ) as a standard Mincer equation.<sup>11</sup> These earnings functions are defined independently of each employment sector by skill level ( $L$ )<sup>12</sup>:

$$\log y_{hi}^{Lj} = X_{hi}^L \Omega^{Lj} + \mu_{hi}^{Lj} \text{ for } i = 1, \dots, n_h \text{ and } j = 2, \dots, J \quad (13)$$

The second component of the total household income, total family *non-labor* income, is the sum of different elements at the household level. This may include international ( $r_h^I$ ) and domestic remittances ( $r_h^D$ ), capital, interest, and dividends ( $k_h$ ), social transfers ( $tr_h$ ), pensions ( $p_h$ ), and other non-labor incomes ( $z_h$ ). Formally,

$$y_{0h} = r_h^I + r_h^D + k_h + tr_h + p_h + z_h \quad (14)$$

From equation (14), attention is given only to modeling international remittances and public transfers while making some minimal assumptions about other components (pensions and capital). In the case of international remittances, migration-related information in most surveys is poor or insufficient, impairs

<sup>11</sup> Mincer (1974).

<sup>12</sup> In this case, a total of 12 Mincer equations should be calculated.

accurate modeling. Instead, the model relies on a simple, non-parametric assignment rule consistent with the existing evidence.

Equations (1) to (14) complete the model. Total household income is a nonlinear function of the observed characteristics of the household and its members and of unobserved characteristics of household members. This function depends on two main sets of parameters: those of the occupational choice model for each skill level; and those in the earning functions for each employment sector and skill level. It is assumed that no variation exists in the composition of the household. In other words, the number, age, and gender of the members of a household remain constant over time. The demographic change is incorporated via calibration of the survey weights. For further details on the estimation strategy of these parameters, see Olivieri, S. et al. (2014).

## Simulation

The second step consists of *replicating* the projected macroeconomic changes (i.e., sector of employment, total output, or public and private transfers) between the baseline and each projected year. These projections derive from various possible changes in different components of the household income generation model (i.e., labor, and non-labor income). This process is divided into three sub-steps ordered in the following sequence: population growth, labor market status and income, and non-labor income.

The population growth adjustment is particularly important in countries with high fertility rates or significant immigration flows, or in cases where the last available national household survey is relatively distant from the projection year. In the first of these instances, the number of labor market entrants rises faster than the overall population. In practical terms, this allows us to explicitly consider changes in the size of the working-age population and hence to distinguish between employment growth driven (or rather absorbed) by demographic trends and net (or additional) employment growth. A simple approach is adopted in this paper to account for population growth with low computational requirement. The estimation weights for all observations are adjusted using neutral distribution to account only for the growth in total population between the baseline survey and the simulated year, maintaining the demographic structure.

One of the transmission channels on which the model focuses is labor markets. This is modeled considering the employment structure and labor earnings projected changes between the baseline and the simulated year. Thus, it allows for changes in employment or earnings or a combination of both. The first stage of the labor market model consists in the allocation of labor status. This step reassigns working-age individuals (i.e., between 15 and 64 years old) between the employment status and across economic sectors by informal and formal status to match the projected aggregate changes in total and sectoral employment. The reallocation method follows Habib et al. (2010).<sup>13</sup>

---

<sup>13</sup> Briefly, the procedure starts with activity and calculates the total number of individuals that need to be reassigned between activity and inactivity statuses. The process continues assigning unemployed and employed population among the actives, and then the employees are allocated across employment sectors to match aggregated changes. The estimated probabilities from the multinomial model are used to select candidates for reassignment in all stages. Error terms are included to represent the unobserved heterogeneity of agents' labor supply behavior. These lead to some disparateness in responses to a change in the

The second sub-step consists of assigning or taking out a labor income to each individual of the working-age population sample according to its "new" labor status. There are three possible cases here. The first case sets positive labor income to zero for those individuals who were employed in the baseline and subsequently become unemployed or inactive as consequence of the macro projection. The second case sees the previous labor income ( $y_{hi}^j$ ) assigned when individuals remain employed in the same employment sector as in the baseline. The third case uses the earnings model estimated as part of the baseline to predict earnings ( $\hat{y}_{hi}^j$ ) for two groups of workers; those with no previous earning history (i.e., those who come from inactivity or unemployment), and those who change employment sector. Formally, the "new" vector of earnings for the working-age population will be defined as:

$$\tilde{y}_{hi}^j = \begin{cases} 0 & \text{when } \hat{U}_{hi}^j \neq \tilde{U}_{hi}^s \text{ for } j>1 \text{ and } s = 0 \text{ or } 1 \\ y_{hi}^j & \text{when } \hat{U}_{hi}^j = \tilde{U}_{hi}^s \text{ for } j=s>1 \\ \hat{y}_{hi}^j & \text{when } \hat{U}_{hi}^j \neq \tilde{U}_{hi}^s \text{ for } j = 0 \text{ or } 1 \text{ and } s>1 \end{cases} \quad \forall i \in [15, 64 \text{ years old}] \quad (15)$$

Where  $\tilde{U}_{hi}^s$  corresponds to the "new" employment status, which corresponds to the maximum of the utility based on the reference status  $s$ . Note that all other workers who do not belong to the working-age population sample are assumed to remain in their baseline employment status as well as receiving their baseline labor earnings.

#### *Matching the total growth*

Once all workers have been assigned positive labor earnings, Olivieri et al. (2014) adjust total earnings in an economic sector to match aggregate projected changes in the sector's output. This adjustment implicitly assumes no differentiation in growth rates between formal and informal earnings in a sector. Then, the authors rescale total earnings once more to account for the change in the economy's total output. The current study accounts for adjusting labor earnings of formal (informal) workers to match projected changes in average formal (informal) labor income of the main activity, given the projected changes in the pseudo-labor productivity of each economic sector, introducing an intra-sectoral variation of labor earnings.<sup>14</sup> This new layer allows more flexibility in the projected labor income distribution and allows incorporating different dynamics by formality status. Then, the model in this paper adjusts labor earnings to match the projected change in average total labor income, given the projected changes in the total pseudo-labor productivity. Finally, total earnings are adjusted using macro projections for economic sectors and total output as in Olivieri et al. (2014). The third sub-step relies then on the fact that projected changes in the sectoral output can be explained by projected changes in sectoral

---

labor demand, capturing the fact that in the real-world individuals who are observationally equivalent (i.e., have identical observable characteristics) might still respond differently to the same change in labor demand - Habib et al. (2010).

<sup>14</sup> Pseudo-labor productivity is the ratio between sectoral GDP, which includes contributions of capital and labor, and total sectoral employment.

employment and projected changes in formal (informal) earnings and profits and assumes that earnings and profits grow at the same rate.<sup>15</sup>

The first step computes the target average labor earnings in the main activity by employment sector ( $\bar{y}_1^{jp}$ ) as the product of the average earnings from microdata at the baseline year ( $\bar{y}_0^{jp}$ ) and the projected growth rate of average labor earnings by employment sector between initial and projected year ( $\delta_{\bar{y}_p}^j$ ). Formally,

$$\bar{y}_1^{jp} = \bar{y}_0^{jp} \left(1 + \delta_{\bar{y}_p}^j\right), \forall j = 2, \dots, J \quad (20)$$

The average labor earnings for the workers in each employment sector is the weighted average of labor earnings in the main occupation in the initial year for all employees in that particular sector:

$$\bar{y}_0^{jp} = \frac{\sum_i (y_{hi}^{jp} w_{hi} \mathbb{1}[\hat{U}_{hi}^j = j])}{N^{jp}}, \forall j = 2, \dots, J \quad (21)$$

where  $N^{jp}$  is total number of employees in the main occupation in the employment sector  $j$ . The second step calculates the "new" average labor earning for workers by employment sector, considering the adjustments already made in labor market and in population growth:

$$\bar{\bar{y}}_1^{jp} = \frac{\sum_i (\tilde{y}_{hi}^{jp} \hat{w}_{hi} \mathbb{1}[\hat{U}_{hi}^j = j])}{\hat{N}^{jp}}, \forall j = 2, \dots, J \quad (22)$$

The third step rescales the "new" average labor income in main and secondary occupations in each employment sector (equation (22)) up to the point where it meets the average labor income target growth (equation (20)) as shown in equation (23), and then rescales labor earnings by the average total labor income for all workers in all sectors. Given that informality status for secondary occupation is missing in most countries, the current method assumes the same status as the main occupation:

$$\bar{y}_1^j = \lambda \bar{\bar{y}}_1^j, \forall j = 2, \dots, J \quad (23)$$

Where  $\lambda$  is the rescaling factor. To replicate the macro-output growth rate by economic sector and the total, the simulation follows Olivieri, S et al. (2014). The output change in each economic sector is apportioned between employment change, earnings change, and adjustments across employment sectors. Given that an individual's labor income depends on his/her employment status and labor earnings, the extent of this change depends on labor and income responsiveness (elasticities) of formal and informal employment in the particular economic sector under consideration. So far, the simulation replicates the average labor income growth rate by employment sector and total. However, since the simulated income growth rate relies on elasticities, it is generally different from that reported by macro projections. Hence, the model prioritizes matching growth rates between macro- and microdata. To do

---

<sup>15</sup> The treatment of public sector workers and those with more than one job follows Olivieri, S. et al. (2014).

that prioritization, first, labor incomes are rescaled, keeping the total volume of the economic activity constant. The result is then shifted by the growth rate of economic activity GDP. The process is then repeated using the total GDP. At the household level, the model also implies that the extent of the impact is dependent on the size of the aggregate change at the economic sector level as well as on the demographics and characteristics of household members, which influence the labor force status and earnings of household members after the change.

To simulate changes in *non-labor* income, projections of changes in public transfers are tailored for each country and year when there is additional information regarding changes in coverage and gratuity; otherwise, they are held constant in real terms. Pensions and capital incomes are assumed to grow at the rate of aggregate GDP for the relevant period, while international remittances follow the methodology of Olivieri et al. (2014). Finally, other non-labor income is assumed to remain constant in real terms.

## Assessment

The final step is the process by which all the information on individual employment status and labor income, together with data on non-labor income at the household level, is used to generate income distributions and to calculate various poverty and distributional measures. These calculations can then be used to compare different scenarios.

### *Assessment of three variations of the model*

This study considers three variations of the microsimulation model to test whether introducing more flexibility in labor incomes leads to more accurate distributional results. In the first variation, the *old model*, the labor income varies only at the economic sectoral and total output growth rates (i.e., agriculture, industry, and services).<sup>16</sup> This structure is less flexible since it considers differentiation by formality status only at the labor market structure level. It imposes sectoral macroeconomic projections on labor income, disregarding within-sector variation on income given by informality. The second structure, named *new model with rescaling*, is the model described above, which incorporates within-sector variation in labor earnings using the formal (informal) projected growth rate in average labor income and then rescales for macroeconomic sectoral and total outputs. Finally, a third and more flexible structure is called the *new model without rescaling*. Like the previous structure, this structure considers projected changes in average labor earnings for each employment sector (formal or informal agriculture, formal or informal industry, formal or informal services), but does not impose macroeconomic changes on income other than those captured through the labor market structure and pseudo-labor productivity.

## Limitations and assumptions

It is important to mention several limitations and assumptions associated with this method, which especially apply when used for projections in the medium/long term. Firstly, the quality of model projections depends on the nature and accuracy of the underlying data. The results are dependent not only on the validity of the micro models but also on the macro projections. The limitations of macro

---

<sup>16</sup> This scenario is based on Olivieri (2020).

projections have been addressed in this study using actual input data. This allows focusing on the predictive capacity of each variation of the proposed micro-simulation method. In addition, using the last available household data as a comparator is tricky because the comparison could potentially attribute specific outcomes to that projection when these outcomes could result from other unrelated factors occurring simultaneously.

Secondly, the simulation relies on behavioral models built on past data that reflect the pre-existing structure of the labor market and household incomes, plus the relationship of these factors and their relationships with demographics as they stood before the expected change. Consequently, the simulation assumes these structural relationships remain constant over the period projections are made. The further back the baseline year is from the present, the more questionable this assumption will likely be.

Thirdly, the model is limited in its ability to account for shifts in relative prices between different sectors of the economy because of external shocks. One such example is the general equilibrium effect of a change in the terms of trade between agriculture and other sectors. In the absence of a CGE model, it is nearly impossible to model changes in terms of trade between economic sectors explicitly.

A fourth consideration is that the model does not consider the geographic mobility of factors (labor or capital) across time. Thus, all individuals are assumed to remain in their place of origin, even as their labor force status changes or their employment sector alters. Usually, this assumption seems like an abstraction from the truth in a stable environment and would only matter when the results are disaggregated spatially or across rural and urban areas.<sup>17</sup>

Fifthly, the simulation component of the model relies on random draw using a pseudorandom number generator in computing the allocations of individuals into labor status, as well as labor earnings of new workers and workers who are changing sectors. The model is configured such that the seed of this random number generator is the same over different runs, so generally, the results will be reproducible. However, small changes in data may lead to changes in outcomes where the random components are required and hence in the results.

Finally, it should be noted that this simulation has not incorporated other transmission channels through which households might be impacted. An example of such impacts is a fall in school retention, educational learning, and childhood nutrition caused by a suspension of school (and school feeding programs).<sup>18</sup>

---

<sup>17</sup> Considering measures against the Covid-19 pandemic, such as curfews, lockdowns, and quarantines, the assumption seems plausible.

<sup>18</sup> See World Bank (2020).

## II. Input data

The best predicting model assessment consists of evaluating which of the previously proposed model variations best capture the observed changes in the income distribution for 15 LAC countries from 2017 to 2020, given the availability of perfect inputs. In other words, this paper attempts to identify the model with the lower bias to estimate changes in the income distribution under the availability of perfect information about the sectoral and total output growth, the total population growth, and the changes in labor market structure and earnings; and to test if allowing for more flexibility in the simulation of labor income results in a more accurate estimation. For this purpose, this paper uses as inputs the World Bank's Macro Poverty Outlooks (MPOs) actual growth in sectoral and total GDP rates and remittances growth rate for the years 2015 – 2020.<sup>19</sup> It also uses the harmonized SEDLAC dataset<sup>20</sup> to compute actual changes between the baseline year and the estimated year in total population growth, labor market structure, and average formal and informal labor income.<sup>21</sup> The simulations performed to test the model's variations use 2016 SEDLAC household survey data for each country (or the most recent household survey data available before 2017) as the baseline for the estimation. It is important to note that formality (informality) has been defined as contributing (not contributing) to work-related retirement insurance for most countries. Table 1 presents the countries considered, the baseline year used for each country, the simulated years, and the informality definition used. It is important to note that, for this work, all inputs are in real terms in 2017 USD PPP, so they already account for inflation changes.

**Table 1 SEDLAC Country Data Used in the Simulations**

Country	Baseline Survey	Simulated Years	Informality Definition
Argentina	2016	2017 – 2020	- Salaried workers who do not receive work-related pension insurance. - Non-salaried workers without complete tertiary education.
Bolivia	2016	2017 – 2020	Workers who do not receive work-related pension insurance.
Brazil	2016	2017 – 2020	- Salaried workers without the work-registry book ("carteira"). - Non-salaried workers who do not contribute to the social security system.

<sup>19</sup> Available at <https://www.worldbank.org/en/publication/macro-poverty-outlook>.

<sup>20</sup> SEDLAC is a database of harmonized socio-economic statistics constructed from Latin American and Caribbean (LAC) household surveys. The SEDLAC database and project were jointly developed and are jointly maintained by CEDLAS (Universidad Nacional de La Plata) and The World Bank's LAC Team for Statistical Development (LAC TSD) in the Poverty and Equity Global Practice. SEDLAC includes information from over 300 household surveys carried out primarily in 18 LAC countries for which a comparable income aggregate (for welfare analysis) can be created: Argentina, Bolivia, Brazil, Colombia, Costa Rica, Chile, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, and Uruguay.

<sup>21</sup> Specifically, the growth rates between the baseline and simulated year are calculated using actual data for participation, informality, employment levels, and earnings by labor sector, inputs which otherwise would be projected using elasticities as in Braga. C., et al. (2023).

Country	Baseline Survey	Simulated Years	Informality Definition
Chile	2015	2017 and 2020	
Colombia	2016	2017 – 2020	
Costa Rica	2016	2017 – 2020	Workers who do not receive work-related pension insurance.
Dominican Republic	2017	2018 – 2020	
Ecuador	2016	2017 – 2020	
El Salvador	2016	2017 – 2019	
Honduras	2016	2017 – 2019	
Mexico	2016	2018 and 2020	Workers who do not receive work-related health insurance benefits.
Panama	2016	2017 – 2019	
Paraguay	2016	2017 – 2020	Workers who do not receive work-related pension insurance.
Peru	2016	2017 – 2020	
Uruguay	2016	2017 – 2020	

Source: Own elaboration based on SEDLAC

Note: Simulated years correspond only to years with the availability of SEDLAC actual data for comparison of the proposed variations of the model.

Following the methodology presented in the previous section, estimation weights are adjusted using neutral distribution for accounting only for the growth in the total population between the baseline survey and the simulated year. The different components of non-labor income follow the methodology described above. However, to facilitate comparisons, international remittances are modeled in all cases using a neutral distribution of the MPOs growth rate for inflows between the base and the simulated years, as follows:

$$r_{1h}^l = r_{0h}^l(1 + \delta_{r^l}) \quad (24)$$

In addition, given that there were several changes in social programs for 2020 due to the COVID-19 pandemic and the lack of macroeconomic projections for these programs, it was necessary to assess the models' goodness of fit using a distribution that excludes those programs. Not all countries included questions in 2020 in their surveys to collect information about the public transfer programs' beneficiaries and amounts. Yet, for the countries where the programs applied during 2020 due to the COVID-19 pandemic were included in the survey and are identifiable, the programs' transferred amount was excluded, and income was re-estimated. This new income is used as the actual distributional estimations. However, households might as well adjust their consumption patterns in response to changes in total household income through changes in their labor market behavior. Hence this new income vector excluding the benefits received from programs, might overestimate the impact of the public transfers programs. Nonetheless, mobility restrictions imposed during the COVID-19 pandemic suggest that this kind of adjustment is limited and that no changes in households' behavior are a reasonable assumption. Table 2 presents the countries and the excluded programs.



### III. Results

This section presents the micro-simulation results for each model variation proposed in the *Assessment* sub-section. The performance of the different variations is measured as their capacity to predict three different aspects of interest: poverty, inequality, and changes along the income distribution. Each aspect's mean squared error (MSE) is calculated to facilitate the analysis. Hence, the best predicting model is the one with the lowest bias compared to the actual value of each measure and for each year. Notice that by using the MSE, top and bottom bias are equally weighted, then the model will be equally poor if it highly overestimates or underestimates these measures. Two time horizons are of interest for the analysis: the short run, which comprises the immediately following two simulated years (2017 and 2018 for most of the countries), and the middle run, which contains the last two simulated years (2019 and 2020 when available).

**Table 2 Public Transfers Programs Excluded for 2020 Assessment, by Country**

Country	Excluded COVID-19 related programs
	No conditional cash transfers programs, this includes:
	- Cash social bonuses
Bolivia	- Transfers from "Familia"
	- Transfers from "Canasta familiar"
	- Transfers from "Universal"
	- Disability bonuses
Brazil	Simulated transfers from "Bolsa Familia"
	- Transfers from "Bono de Emergencia COVID-19"
Chile	- Transfers from "Ingreso Familiar de Emergencia"
	- Bonus "Ayuda Familiar"
	- Transfers from "Proteger"
Costa Rica	- Cash transfers due to COVID-19 pandemic
	- Non-cash transfers due to the pandemic
	- Transfers from "Plan Quedate en Casa"
Dominican Republic	- Transfers from "FASE – Fondo de Asistencia Solidaria al Empleado"
	- Transfers from "Asistencia al Trabajador Independiente"
Paraguay	- Transfers from Tekopora
	- Transfers from Pytyvo
	No conditional cash transfers programs, this includes:
	- Transfers from "Bono Yo me quedo en casa"
Peru	- Transfers from "Bono Independiente"
	- Transfers from "Bono Rural"
	- Transfers from "Bono Familia"

Source: Own elaboration based on SEDLAC.

## Poverty

This section compares models' predicting performance using international poverty and vulnerability thresholds. More specifically, it shows how well the model fits in estimating the headcount rate at the three international poverty lines (2.15, 3.65, and 6.85 USD), and the vulnerability (6.85 – 14 USD), middle-class (14 – 81 USD), and upper-class thresholds (> 81 USD). Hence, this exercise captures changes in the whole distribution using reduced cut-offs (the international comparison parameters) as a reference. The deviation and square error from actual headcounts are computed for measuring each year and each model variation's performance. The overall performance for each model variation is estimated as the average of squared errors (Mean Squared Error – MSE) across countries and headcounts by year. Formally,

$$\overline{MSE}_t = \frac{1}{N} \sum_c^C \sum_m^M MSE_t^{c,m} \quad (25)$$

Where:

$c$  = country

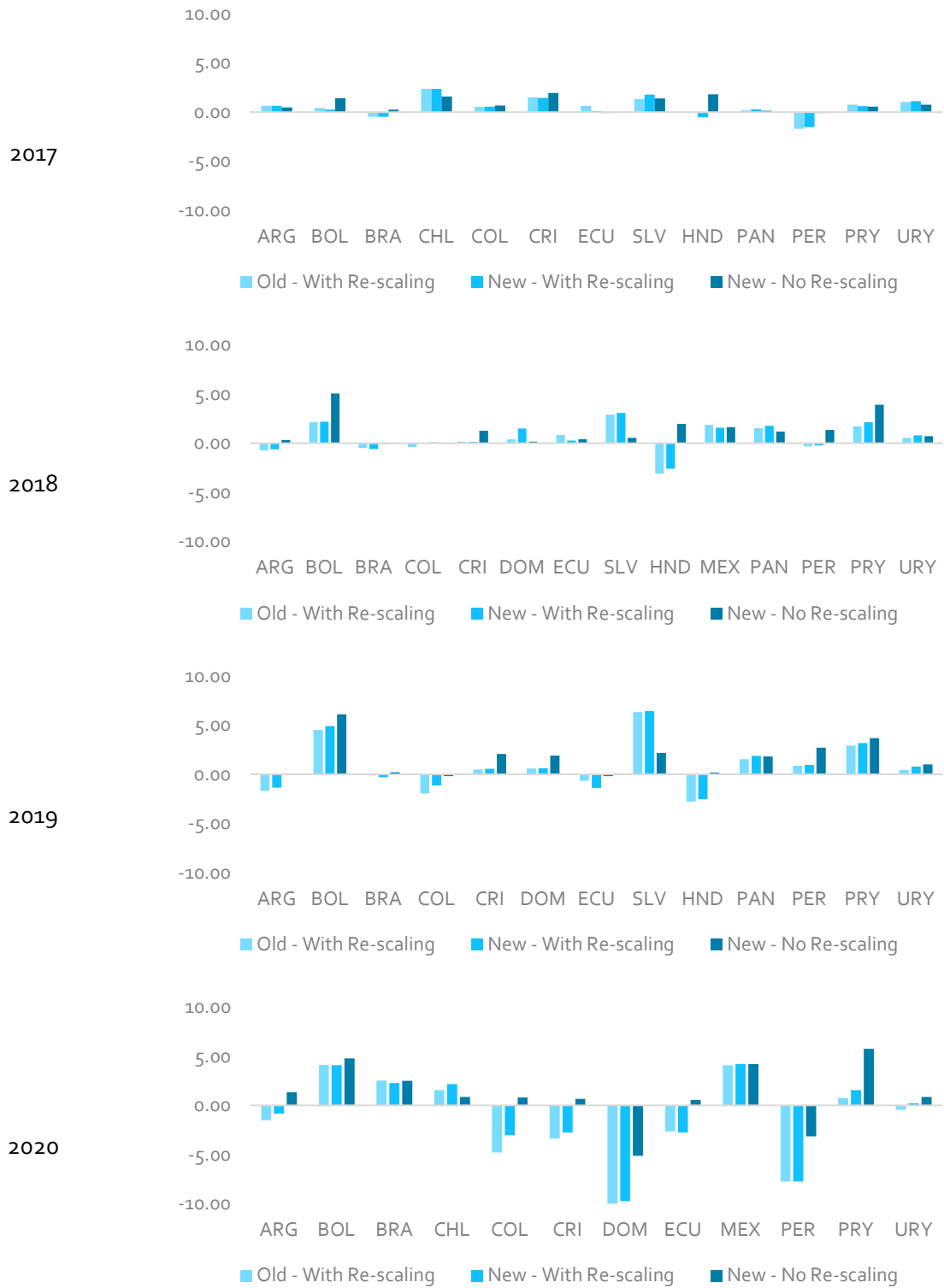
$m$  = measure (i.e., poverty at the three international lines, vulnerability, middle class, and upper class)

$N$  = number of measures considered (i.e., six measures in this case)

Results indicate that the three variations of the model generally have outstanding performances but do much better in the short run (2017 and 2018), where the different simulation models provide very similar estimations irrespective of the country's poverty level (see Figure 1 for USD 6.85 a day in 2017 PPP). Further, in 2017, the *old model* and the *new model with rescaling* underestimate moderate poverty in some countries. This result is exacerbated in the middle run (see 2019 and 2020) since errors increase across the model's variations for the upper international poverty line. The differences in the medium run go hand-in-hand with cross-country variations, making it challenging to identify the cross-country best-predicting variation of the model. For instance, in countries such as Bolivia and Paraguay, the less flexible variations outperform the *new model without rescaling*, while the latter seems to do better for all other countries.

On the other hand, poverty MSEs are significantly low for all the model's variations in the short-run, making the best-predicting model change depending on the country and the simulated year (Figure 2). These results suggest that, independently of the labor income modeling approach, the proposed microsimulation methodology is very good at forecasting poverty in the short run. However, the *new model without rescaling* (the most flexible variation) is the best to capture changes at different thresholds of the income distribution in the medium run, even if 2018 and 2019 are considered instead of 2019 and 2020. This way, the *new model without rescaling* stands out as the best option to estimate poverty in LAC countries. In addition, results show an increasing error trend when the time interval widens from the baseline year or when analyzing atypical years such as 2020. Nonetheless, even in exogenous shocks such as the COVID-19 pandemic, the models' estimations are relatively close to actual values, as shown in Figure 1.

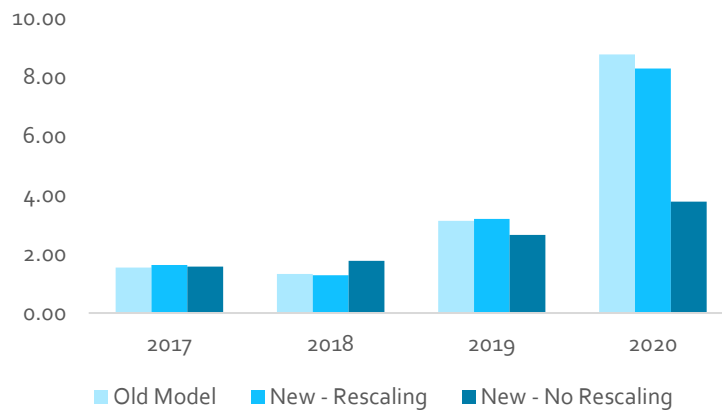
Figure 1 Poverty Incidence at USD 6.85 a day 2017 PPP for LAC Countries, Actual vs. Projected



Source: Own estimations based on SEDLAC.

Note: Actual data excludes COVID-19-related mitigation measures when the survey includes the program beneficiaries and the amount received.

**Figure 2 Average MSE of Poverty Measures for LAC Countries**



Source: Own estimations based on SEDLAC.

The difference between actual and estimated poverty gap values is presented in Figure 3. The poverty gap measures the ratio between the shortfall of the household per capita income from the poverty line and the poverty line. In general, results suggest that in the pre-COVID-19 pandemic scenario, the different 'model's variations overestimate the poverty gap. However, the magnitude varies once again with the country. Countries such as Bolivia and El Salvador have higher levels of overestimation in the poverty gap, while the measure is underestimated in Brazil.

Like in the poverty incidence case, the MSE for the poverty gap indicates that all the proposed models' variations predict better in the short run with very little difference between the *old model* and the *new model with rescaling* (Figure 4). This means the variations succeed at identifying the poor and estimating the intensity of poverty in the most immediate years. Yet, the performance of the *new model without rescaling* is poorer. However, in the medium run, the less flexible variations (*old model* and *new model with rescaling*) fall behind *the new model with rescaling* since the trend of their error increases exponentially over time. This way, both poverty headcount and the poverty gap indices suggest that, in the short run, there is little difference among the analyzed model variations. Even the simplest one (the *old model*) brings reliable results. Still, in the middle run, it is necessary to incorporate more flexibility in the income structure to capture the changes in poverty and vulnerability measures. All of this is subject to country-specificities, as shown before.

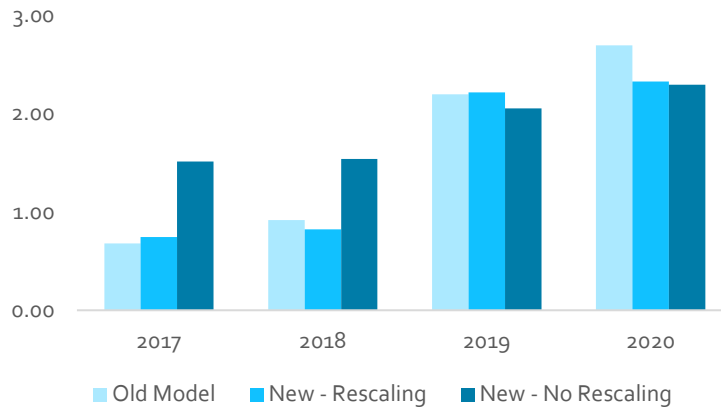
**Figure 3 Difference Between Actual and Projected Poverty Gap for LAC Countries, by Model**



Source: Own estimations based on SEDLAC.

Note: actual data excludes COVID-19- related mitigation measures when the survey includes the program beneficiaries and the amount received.

**Figure 4 Average MSE of Poverty Gap at 6.85 USD 2017 PPP, for LAC Countries**



Source: Own estimations based on SEDLAC.

## Inequality

This section presents results for the three model variations when estimating the most standard inequality measure – the Gini coefficient.<sup>22</sup> Results indicate that, like the case of the international upper poverty line, the prediction is similar for the three model variations in the short run, and the fit is generally reasonable (Figure 5). Still, errors increase with the distance between the baseline and the estimated year. In addition, results show that the assessed variations tend to constantly overestimate inequality for some countries (i.e., Bolivia, Costa Rica, El Salvador, Honduras, Mexico, Panama, Peru, and Paraguay) and underestimate it in others (like the Dominican Republic). Thus, the bias is not always in the same direction or magnitude. For instance, the *new model with rescaling* has the lowest gap in Bolivia in 2017 but the largest from 2018 through 2020.

Figure 6 presents the MSEs for LAC countries by year and its trend to compare all the model variations over time. In this case, errors seem to increase in a linear pattern instead of the exponential-shaped trend observed in the poverty and vulnerability indicators. Overall, results suggest that the *old model* is the best variation of the micro-simulation model to estimate inequality since its MSE is constantly lower than the other tested variations, except for atypical years such as 2020.

In summary, results suggest that no variation of the model should be considered a best-all-countries/all-measures fit since country-specificities might arise. Depending on the purpose of the analysis, either the *new model without rescaling* (poverty and vulnerability) or the *old model* (inequality) might perform better. Then, to choose a model, it is necessary to deepen the analysis to obtain a variation that works decently for both poverty and inequality measures. To this end, an additional assessment of the models is made to check their performance in estimating changes along the whole income distribution. The results obtained for this evaluation are presented in the following section.

<sup>22</sup> This indicator measures how dispersed the income distribution is and takes values from 0 to 1, where 0 represents perfect equality and 1 stands for perfect inequality.

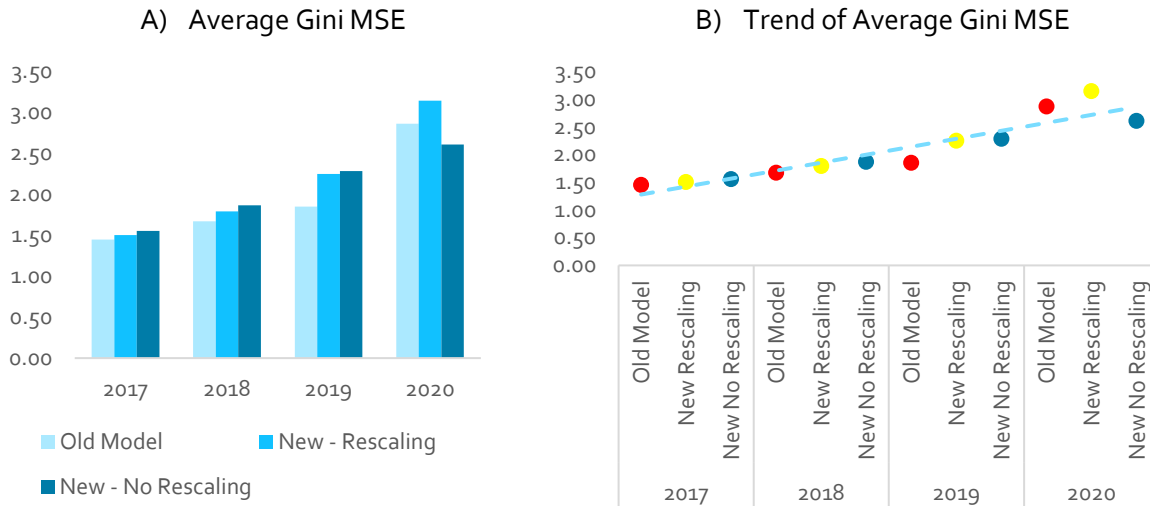
**Figure 5 Gini Coefficient for LAC Countries, Actual vs. Projected**



Source: Own estimations based on SEDLAC.

Note: actual data excludes COVID-19-related mitigation measures when the survey includes the program beneficiaries and the amount received.

**Figure 6 Average MSE and Trend of Gini, for LAC Countries**

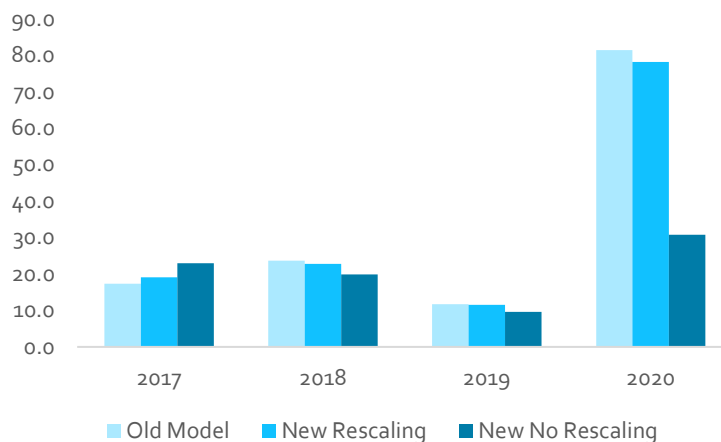


Source: Own estimations based on SEDLAC.

### Distributional effects

This section presents the performance of the considered models' variations along the whole income distribution. The Growth Incidence Curves (GICs) for each country/year are calculated using the SEDLAC actual income data and the simulated income vector for each variation. The squared errors correspond to the difference in the annualized growth rate at each percentile of the income distribution. The MSEs correspond to the average squared error (MSE) across countries over time. Figure 7 presents the results. It is worth noting that these results do not include information on the 1-5 and 96-100 percentiles due to the high dispersion in the tails of the actual distribution.

**Figure 7 Average MSE of GICs for LAC Countries**



Source: Own estimations based on SEDLAC.

Results show that differences between the proposed variations of the model are more significant in this case (a higher magnitude in the units of the vertical axis), and that is because more cut-offs are



considered in the analysis (i.e., each income percentile). Except for the immediately following year to the baseline, the *new model without rescaling* performs better overall than the other variations of the micro-simulation model. The difference is prominent in atypical years like 2020. In addition, there is not a clear trend in the errors like in the previous cases since the errors have an inverted-U shape - a decrease in the first years and then a remarkable increase-. However, the exponential growth in the presence of shocks is salient for the less flexible variations.

At the country level, Table 3 presents the share of percentiles that fall inside the actual GIC Confidence Interval of 95% for each model's variation. This is the number of percentiles with an estimated growth rate that falls inside the confidence interval of the respective actual growth rate, divided by the total number of percentiles. In the table, columns with numbers "1", "2", and "3" contain the share of percentiles within the confidence interval for the *old model*, the *new model with rescaling*, and the *new model without rescaling*, respectively. In the best case, all percentiles fall inside the confidence interval of 95%, which means a share of 100. Results indicate that there is still high variation across countries and years. Yet, the *new model without rescaling* estimates the income growth along the distribution more accurately since the average share for all countries is higher for all the years included in the analysis.

#### *Good fit of the models' variations*

In addition, Table 3 shows cases where the model's variations, especially the most flexible one, perform exceptionally well (shares over 80 for the *new model without rescaling*), such as Bolivia 2019 and Dominican Republic 2018. In these cases, the estimated growth of income along the whole distribution is very close to the actual growth for the *new model without rescaling*, as shown in Figure 8. In contrast, the less flexible variations are more distant from the actual distribution and even out of the confidence interval, as in the case of Bolivia 2019. The latter is an interesting case since, as shown before, Bolivia is one of the countries with the poorest fit in poverty estimation. The explanation for this discrepancy relies on the measure of analysis. Poverty measures the number of people whose income falls below a determined poverty line. In this sense, this measure depends on the selected threshold and the income growth for the population most likely to have income below that threshold. In the case of Bolivia, the GICs presented in Figure 8 show a good level of adjustment overall but insufficient income growth below the 31 percentile, where the poverty lines probably fall. This example highlights the importance of considering the models concerning the target measure and the measured object of analysis (e.g., poverty, inequality, income distribution).

#### *Divergence of distribution tails*

In other cases, the fitting is not so good along the whole distribution. Argentina 2019 and Panama 2019 are examples of these cases (Figure 9). Notably, these countries show that the three assessed model variations might have difficulties estimating income variation in the distribution's tails. A possible explanation might be due to the limited employment sector classification. The micro-simulation model relies on a 6-sectors classification (formal agriculture, informal agriculture, formal industry, informal industry, formal services, and informal services), and movements of labor earnings according to aggregate output or average earnings are not enough to capture intra-sectoral variations. That said,

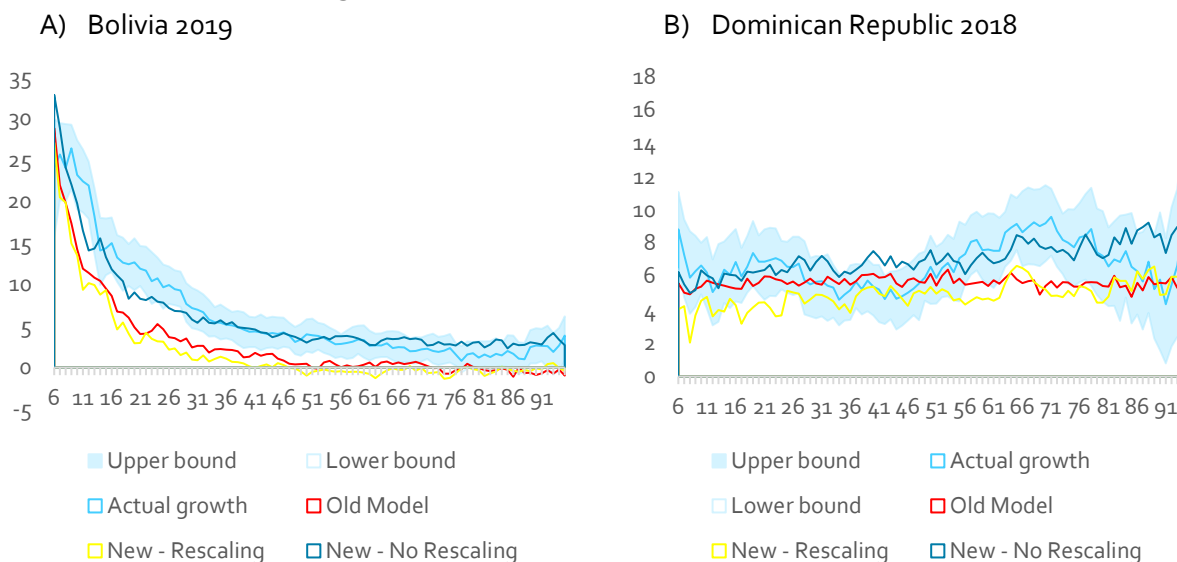
analyzing this possibility is beyond the scope of the current analysis. Nonetheless, in both cases presented in Figure 9, the *new model without rescaling* outperforms the other proposed variations of the model, indicating that the *new model without rescaling* is very good at estimating changes along the income distribution with limited intra-sectoral information as input.

**Table 3 Share of Percentiles that Fall Inside the actual GIC Confidence Interval , by Country and Year**

Country	2017			2018			2019			2020			Average		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
ARG	5	5	12	0	0	64	29	34	47	19	24	55	13	16	44
BOL	53	54	39	21	20	17	14	18	85	24	21	34	28	28	44
BRA	44	44	38	30	29	52	29	30	30	24	24	13	32	32	33
CHL	0	0	0							11	18	28	6	9	14
COL	24	23	40	14	16	17	7	9	61	2	8	30	12	14	37
CRI	35	38	36	43	47	76	42	46	82	3	5	45	31	34	60
DOM				74	55	82	96	100	4	8	13	57	59	56	47
ECU	26	39	56	67	60	83	60	59	86	5	5	53	39	41	69
SLV	25	20	23	15	18	36	2	1	3				14	13	21
HND	16	13	31	25	33	32	56	51	15				32	32	26
MEX				23	26	20				7	6	6	15	16	13
PAN	45	52	59	7	7	7	81	86	77				44	48	47
PRY	65	58	71	33	38	7	18	25	53	12	12	45	32	33	44
PER	8	11	74	46	47	26	28	28	26	4	5	5	22	23	33
URY	5	5	45	4	4	72	9	11	73	5	4	88	6	6	69
<b>Average</b>	<b>27</b>	<b>28</b>	<b>40</b>	<b>29</b>	<b>29</b>	<b>42</b>	<b>36</b>	<b>38</b>	<b>49</b>	<b>10</b>	<b>12</b>	<b>38</b>	<b>26</b>	<b>27</b>	<b>40</b>

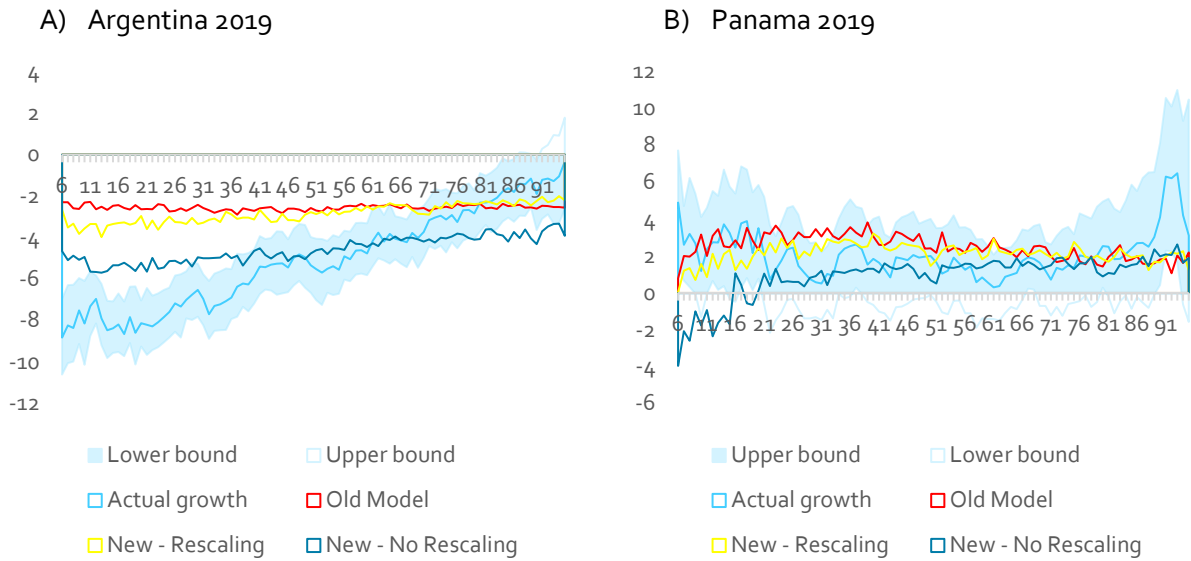
Note: 1 Corresponds to the *old model*, 2 to the *new model with rescaling*, and 3 to the *new model without rescaling*.  
Source: Own elaboration based on SEDLAC

**Figure 8 Examples of Countries With Good Fit**



Source: Own estimations based on SEDLAC.

**Figure 9 Examples of Countries With Inferior fit in the Distribution Tails**

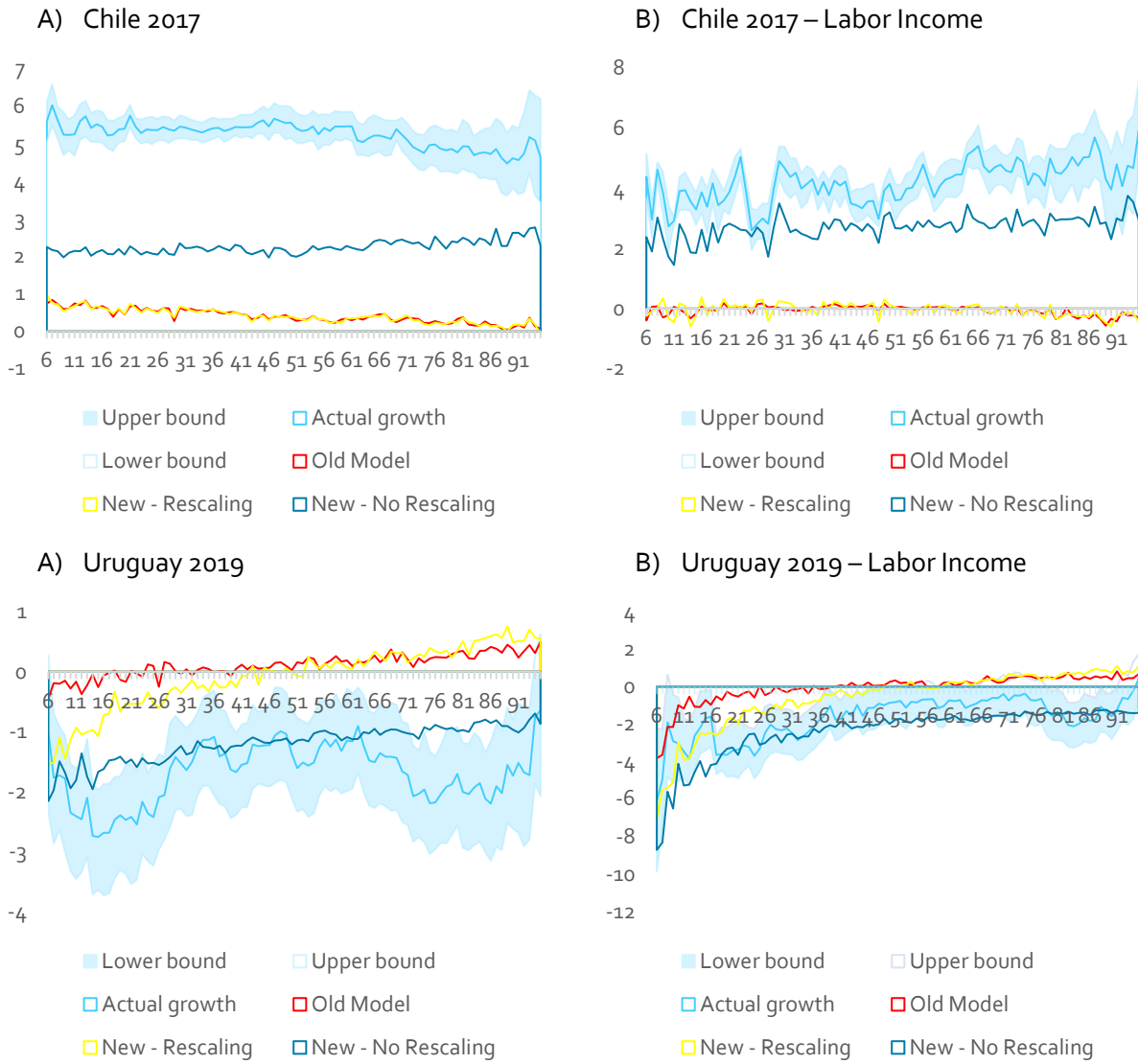


Source: Own estimations based on SEDLAC.

*Scale and flexibility factors*

Table 3 also shows cases where the percentage of percentiles that fall inside the confidence interval is very small or even zero. The GICs for labor and total income are presented for Chile 2017 and Uruguay 2019 in Figure 10. In the case of Chile 2017, the GICs show a scale factor indicating that the estimated income growth falls short when compared to the actual growth at each distribution percentile. For Uruguay 2019, the GICs show that there are some parts of the income distribution where more flexibility is necessary, and they do not necessarily correspond to the tails of the distribution (percentiles 13-20 and 71-88). In both cases, the GICs for only labor income indicate that the *new model without rescaling* estimates an income growth rate closer to the actual growth than the other models. This way, results suggest that the differences in levels are mainly due to the non-labor income, where more flexibility might contribute to better estimating the changes in the total income distribution. Some refinements of this methodology could help improve the estimation of non-labor income and the overall estimate, like modeling public transfers or applying different assumptions on remittances, capital, and pensions. In an exercise like this paper, Cojocarú and Olivieri (2014) find that accounting for social protection benefits in a micro-simulation model applied to Serbia 2009 reduced the bias in poverty estimates statistically indistinguishable from the actual headcount. An applied example of how refinements in non-labor income modeling, more specifically remittances, affect the estimation results is presented in **Box 1** for El Salvador.

**Figure 10 Examples of Countries With Inferior Fit Due to Non-Labor Income**



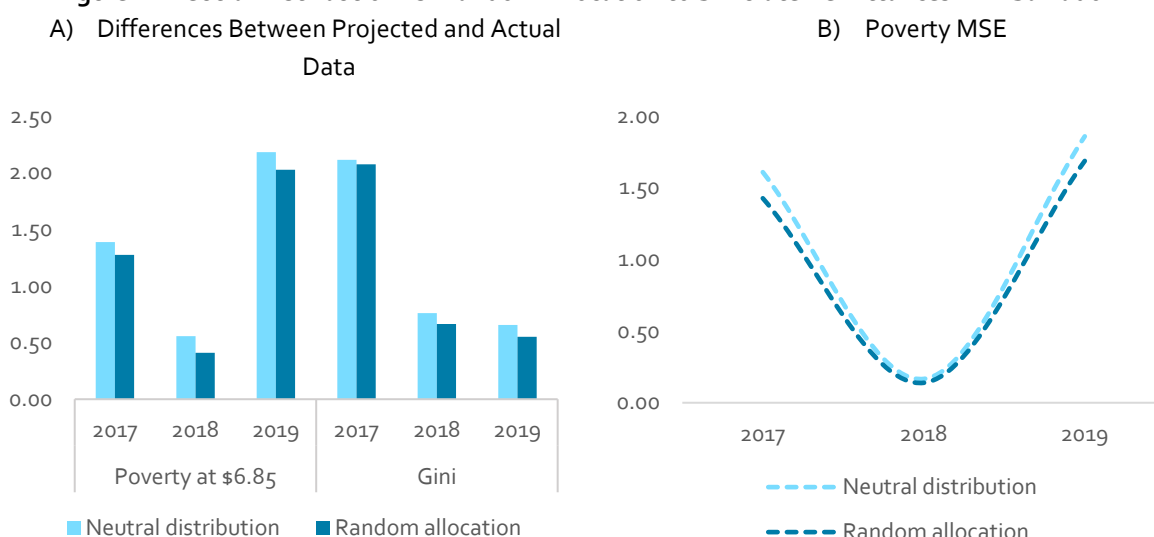
Source: Own estimations based on SEDLAC.

## Box 1: Effects of Changing Remittances Modeling in El Salvador

Some refinements on how non-labor income is projected may result in a more accurate estimation of the changes along the income distribution. This study modeled international remittances as neutral distribution using the inflows growth rate between the baseline year and the simulated year from the World Bank's Macro Poverty Outlooks. However, as stated in the methodology presented earlier in this paper, international remittances can also be modeled using the proposed method in Olivieri et al. (2014). This methodology consists of a two-step assignment rule. First, the projected amount of international remittances is calculated as the initial level of remittances times the change between the baseline and the projected year; second, the difference in the amounts between the base year and simulated year is randomly assigned to households within each region, considering population growth while maintaining the original regional distribution of international remittances.

To assess the effect of these two different international remittance approaches on income distribution, this document simulates the poverty and vulnerability headcounts and the Gini coefficient for El Salvador 2017 - 2019. It is important to clarify that El Salvador is selected due to its high level of international remittances. In 2019, this non-labor income component accounts for 61.6% of the average per-capita non-labor income of the country<sup>23</sup>; hence, a change in the way international remittances is modeled is expected to impact the overall performance of the model. **Figure 11** presents the results for this exercise using the *new model without rescaling*. In the results, Neutral distribution corresponds to the first approach while Random Allocation refers to the methodology in Olivieri, et al. (2014).

**Figure 11 Neutral Distribution vs. Random Allocation to Simulate Remittances in El Salvador**



Source: Own estimations based on SEDLAC.

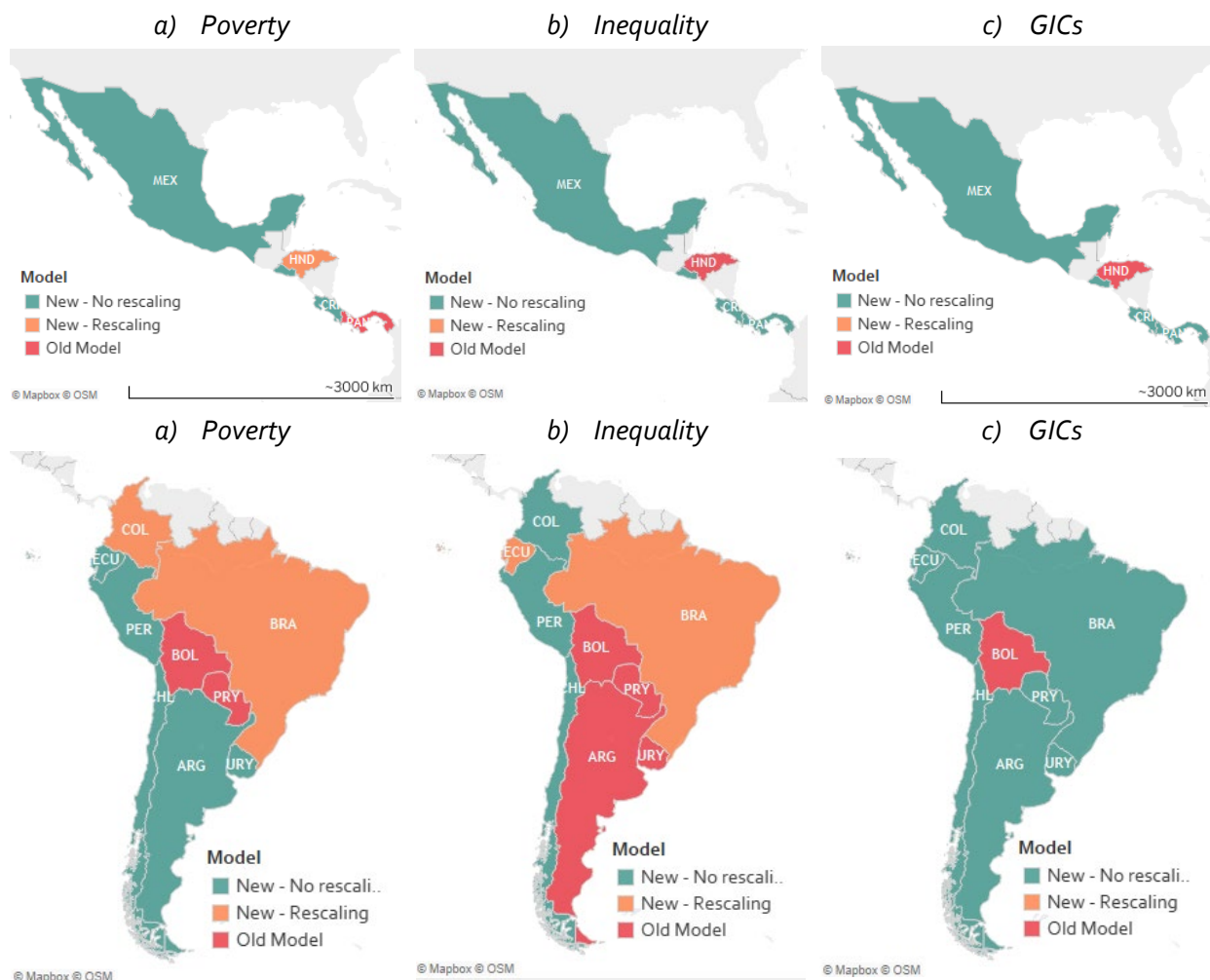
As expected in a country like El Salvador, results suggest that using random allocation to model international remittances slightly improves the model's performance along the whole time series compared to neutral distribution. The difference between the simulated and actual poverty rate at 6.85 USD reduces for all years by around 0.14 percentage points (p.p.) (Figure 11.A). Still, the reduction in inequality is smaller (approximately 0.08 p.p.). In addition, results suggest that the average MSE for poverty and vulnerability is also lower with the random allocation approach (Figure 11.B), especially in years where the model's bias is larger such as 2017 and 2019.

<sup>23</sup> Own calculations using SEDLAC 2019.

## The best predicting model variation

Overall, results suggest that incorporating more flexibility in the labor income modeling translates into smaller bias across time and a more accurate estimation of the changes in the income distribution. In summary, Figure 12 shows the model's variation that, on average, has a better fit (lower MSE) across the four years in the different welfare measures analyzed by country. Notice that the new and more flexible variation outperforms the other two tested variations in several countries for poverty and inequality and in almost all countries when using changes in the entire income distribution as the selection criterion. Bolivia and Honduras are the only countries where it seems better to implement the less flexible version (the *old model*) of the micro-simulation model.

**Figure 12 Best Model Predicting Poverty, Inequality, and the Income Distribution by Country**



Source: Own estimations based on SEDLAC.

## IV. Final remarks

This paper assesses the predictive capacity of specific microsimulation models using a minimum set of perfect macroeconomic input data and under different assumptions on how family income is estimated. Using actual input data allows for isolating the analysis from any possible bias from macroeconomic inputs, so the predictive capacity of the model variations relies only on the changes in the assumptions made in the proposed microsimulation methodology. The study makes two contributions: first, it extends previous microsimulation methodologies by accounting for labor income movements within each economic sector given by informality, and second, it tests three versions of the model (i.e., the old model, the new model with rescaling, and the new model without rescaling) to identify the one that fits better the actual income distribution. The *old model* corresponds to the first option in which labor income varies only at the macroeconomic sectoral and total output growth rates; the *new model with rescaling* is the second, which incorporates within-sector variation in labor earnings using the formal (informal) changes in average labor income in the main activity, provided by the labor income-pseudo-labor productivity elasticities for each employment sector, and rescales for sectoral and total output growth. The last option is the *new model without rescaling*, which only considers changes in average income earnings given by changes in average labor earnings by employment sector. The proposed alternatives are tested using the SEDLAC harmonized household data for 2016 or the closest year to predict poverty, inequality, and the GICs in the short (2017 – 2018) and medium run (2019 – 2020) for 15 LAC countries.

Overall, results suggest that incorporating more flexibility in the labor income estimation by adding labor market attributes such as informality produces a smaller bias across time and a better estimation of the changes along the income distribution. For poverty, results indicate that the three model variations generally have outstanding performances in the short run, irrespective of the country's poverty level. Hence, in the short run, these alternative models do not differ largely in performance gains; even the *old model* brings reliable results. However, the *new model without rescaling* is the best at identifying poor people in the medium run, meaning that accounting for labor market aspects such as informality produces more accurate estimates in contexts of high uncertainty. On the other hand, results indicate that the *old model* is best at estimating inequality, except for atypical years like 2020, but the advantage over the other model variations is slight. Moreover, the assessed model's variations overestimate inequality for several countries (i.e., Bolivia, Costa Rica, El Salvador, Honduras, Mexico, Panama, Peru, and Paraguay) and underestimate it in a few others (like the Dominican Republic). Thus, results might diverge, and the same micro-simulation model option is not be the best for poverty and inequality.

This document extends the analysis along the income distribution using GICs. In this case, the *new model without rescaling* outperforms the other two options, and the difference is significant in atypical years like 2020. Therefore, introducing intra-sectoral variation through differentiated growth by informality status helps translate macroeconomic movements into more precise growth at every level of the income distribution. Three lessons can be drawn from this exercise: first, when the proposed micro-simulation model performs exceptionally well, the *new model without rescaling* variation is remarkably superior to the others. Second, all variations find it challenging to estimate income changes in the tails, possibly due to limited variation in the labor market setting that does not allow to account for other specificities

besides informality. Yet, *the new model without rescaling* outperforms the others in these situations. Third, when there is a big gap in growth (i.e., the estimated growth falls outside the confidence interval of the actual growth), results suggest that differences are mainly due to insufficient growth in non-labor income. In summary, adding more flexibility in the labor income modeling contributes to a better estimation of the changes in the income distribution.

Finally, evidence suggests that refinements on the non-labor income side improve the model performance. In particular, changing how international remittances are simulated from neutral distribution to random allocation for countries with a high share of inflows, such as El Salvador, resulted in a slightly lower bias both in poverty and inequality measures.



## References

- Braga, C., Montoya, K., and Olivieri, S. (2023). Forecasting labor market dynamics in LAC with minimum information: an elasticity approach. World Bank – *forthcoming*.
- Bourguignon, F., and Spadaro, A. (2006) "Microsimulation as a Tool for Evaluating Redistribution Policies." *Journal of Economic Inequality* 4 (1): 77–106.
- Bourguignon, F. and Ferreira, F. (2005). "Decomposing Changes in the Distribution of Household Incomes: Methodological Aspects" in *The microeconomics of income distribution dynamics in East Asia and Latin America*, eds. Bourguignon, F., Ferreira, F., and Lustig, N. Washington, DC: Oxford University Press and The World Bank.
- Bourguignon, F., Bussolo, M. and Pereira da Silva, L. (2008). "The impact of macroeconomic policies on poverty and income distribution: macro-micro evaluation techniques and tools" in *The Impact of Macroeconomic Policies on poverty and Income Distribution*, eds. Bourguignon, F., Bussolo, M., and Pereira da Silva, L. The World Bank, Washington, DC.
- Bourguignon, F., Robilliard, A.S., and Robinson, S. (2003) "Representative versus real households in the macro-economic modelling of inequality", DT/2003/10, DIAL/Unite de recherche CIPRE.
- Buddelmeyer, H., Hérault, N., Kalb, G., and van Zijll de Jong, M. (2008) "Disaggregation of CGE results into household level results through micro-macro linkage: Analysing climate change mitigation policies from 2005 to 2030", Melbourne Institute Report No 9, Melbourne Institute of Applied and Economic Research, University of Melbourne.
- Cojocar, A., and Olivieri, S. (2014). "Updating the Poverty Estimates in Serbia in the Absence of Micro Data: A Microsimulation Approach". Policy Research Working Paper, no. WPS 6889. Washington, D.C.: World Bank Group.
- Colombo, G. (2008). Linking CGE and microsimulation models: A comparison of different approaches. None.
- Ferreira, F., Leite, P., Pereira da Silva, L., and Picchetti, P. (2008). "Can the Distributional Impacts of Macroeconomic Shocks Be Predicted? A Comparison on Top-Down Macro-Micro Models with Historical Data for Brazil" in *The impact of macroeconomic policies on poverty and income distribution: macro-micro evaluation techniques and tools*, eds. Bourguignon, F., M. Bussolo, and Pereira da Silva, L. The World Bank, Washington, DC.
- Ferreira, J., and Horridge, M. (2005). The Doha round, poverty and regional inequality in Brazil. World Bank Policy Research Working Paper, (3701).
- Habib, B., Narayan, A., Olivieri, S. and Sanchez-Paramo, C (2010) Assessing Poverty and Distributional Impacts of the Global Crisis in the Philippines. A microsimulation approach, PRWP 5286, Washington DC
- Hérault, N. (2010). Sequential linking of computable general equilibrium and microsimulation models: a comparison of behavioural and reweighting techniques. *International Journal of Microsimulation*, 3(1), 35-42.

- MacFadden, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior" in *Frontiers in Econometrics*, ed. Zarembka, P. New York, Academic Press.
- Mincer, J. (1974) "Schooling, experience and earnings", New York: Columbia University Press for NBER.
- Olivieri, S., Kolenikov, S., Radyakin, S., Lokshin, M., Narayan, A., and Sanchez-Paramo, C. (2014), "Simulating Distributional Impacts of Macro-Dynamics: Theory and applications", The World Bank, Washington, DC.
- Olivieri, S. (2020) "Pincered: the welfare and distributional impacts of the 2020 triple crisis in Ecuador," Technical Report, Manuscript World Bank April 2020.
- Savard, L. (2003) "Poverty and Income Distribution in A CGE-Household Micro-Simulation Model: Top-Down/Bottom-Up Approach", CIRPEE Working Paper 03-43.
- Train, K., and Wilson, W. (2008). "Estimation on Stated-Preference Experiments Constructed from Revealed-Preference Choices." *Transportation Research Part B: Methodological* 42 (3): 191–203.
- World Bank (2020). *Macro-Poverty Outlook*, Washington, DC.