TECHNICAL NOTE Nº IDB-TN-2905

Nowcasting Poverty in Central America, Panama, and the Dominican Republic: A Micro-simulation Approach.

Carlos Eggers
Lucía Martín
Laucel Muñoz
Álvaro Salamanca

Inter-American Development Bank
Country Department Central America, Haiti, Mexico, Panama and the Dominican Republic

April 2024

Nowcasting Poverty in Central America, Panama, and the Dominican Republic: A Micro-simulation Approach.

Carlos Eggers
Lucía Martín
Laucel Muñoz
Álvaro Salamanca

Inter-American Development Bank
Country Department Central America, Haiti, Mexico, Panama and the Dominican Republic

April 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Nowcasting poverty in Central America, Panama, and the Dominican Republic: a micro-simulation approach / Carlos Eggers, Lucía Martín, Laucel Muñoz, Álvaro Salamanca.
p. cm. — (IDB Technical Note ; 2905)
Includes bibliographic references.
1. Poverty-Central America-Data processing. 2. Poverty-Panama-Data processing. 3. Poverty- Dominican Republic-Data processing. 4. Poverty-Econometric models-Central America. 5. Poverty-Econometric models-Panama. 6. Poverty-Econometric models-Dominican Republic. I. Eggers Prieto, Carlos. II. Martín Rivero, Lucía. III. Muñoz, Laucel. IV. Salamanca, Álvaro. V. Inter-American Development Bank. Country Department Central America, Haiti, Mexico, Panama, and the Dominican Republic. VI. Series.
IDB-TN-2905

Keywords: poverty, nowcasting, household surveys, Central America.
JEL Codes: C53, D31, I32, O10

http://www.iadb.org

Nowcasting poverty in Central America, Panama, and the Dominican Republic: a micro-simulation approach*

Carlos Eggers† Lucía Martín‡ Laucel Muñoz § Álvaro Salamanca¶

April, 2024

Abstract

Official poverty rates are derived from household surveys that often have limited frequency, unexpected gaps due to fieldwork constraints, and substantial delays in processing and publication. This paper presents a novel micro-simulation method for estimating poverty, which introduces changes in demographic and labor variables into the surveys. These changes rely on a few observed or forecasted standard macroeconomic indicators, making our method easy to replicate across countries and in different years, which is not the case for the majority of other micro-simulation techniques. We present an application to the case of Central America, Panama, and the Dominican Republic (CAPDR) and show that it outperforms the fit of other methods that solely rely on direct imputations from GDP to households' income. We argue that the simulation of changes in unemployment, and not only income, could also capture more precisely the impact of large economic fluctuations on poverty rates. Finally, although our method proves to work by simulating changes only in labor income, it still leaves room to simulate variations in other sources of income.

Keywords: poverty, nowcasting, household surveys, Central America.

JEL Codes: C53, D31, I32, O10

*We thank the team of economists of the Central America, Haiti, Mexico, Panama and the Dominican Republic department at the Inter-American Development Bank for their useful comments and suggestions.

†Boston College: eggersc@bc.edu

‡Inter-American Development Bank: lmartin@iadb.org

§DMA Research and Data Analytics: laucel@dmaanalytics.com

¶DMA Research and Data Analytics: alvaro@dmaanalytics.com

1 Introduction

Monitoring poverty rates in developing countries is a crucial task to foster well targeted policies, and execute efficient budget designs in terms of welfare. Also, policy institutions that work for development need to have reliable tracking of the social conditions in different countries, in order to prioritize projects both between and within them. Poverty rates and other social indicators are typically measured using household surveys, which are not always available at the moment in which many policy decisions are made. For that reason, a reliable method to *nowcast* poverty is highly convenient to guide social policy, especially in countries that have unstable or very high growing macroeconomic indicators, being exposed thereby to large changes in the poverty rate.

As poverty is defined by households' income, which is in turn the sum of many income sources that change for reasons that are idiosyncratic for each country, period and type of household, the most precise way to predict changes in poverty consists on micro-simulations that modify the last observed survey in a way that is consistent with macroeconomic variables, fiscal policies and international flows. Typically, this is made when specific events with special characteristics occur, and the social effects of those events need to be estimated. For example, during the 2020 COVID-19 crisis, many micro-simulations were done to predict the poverty rate increase due to the crisis. Some articles where those methods were applied for Latin America are López et al. (2020), Acevedo et al. (2020), and Brum and De Rosa (2021). However, designing micro-simulations that take full acknowledgment of idiosyncratic events of each country and period, is a procedure that, although is likely to be more precise for the case it is focusing on, it is also likely to be impossible to replicate exactly for other countries, or even other periods of the same country. For that reason, maximizing precision of a nowcasting procedure, comes in all likelihood at the cost of being very hard to track each year, and unmanageable to make for a large set of countries.

The need of replicability has been a motivation for some other methods applied in groups of countries for which additional data and tools are available. A notable example that uses microsimulations is the nowcasting of poverty in Europe described in Navicke et al. (2014). In the case of Central America, Panama, and the Dominican Republic (CAPDR) however, that kind of data and tools are not available. A method that has high level of replicability, and can be used in almost any country, is the one proposed by Mahler et al. (2022). The authors show how a simple method that scales the distribution of income in proportion to the predicted change of the mean. Those changes are typically predicted using GDP per capita growth at PPP, but other variables such as consumption can be used in some cases to improve the performance of the model. This method gives fairly precise estimates of poverty rates at global level, using just GDP per capita as predictor of changes in the mean of the distribution. However, this approach lacks reliability when it is applied in some individual countries, because of two reasons:

(1.) Even if on average, poverty rates are driven almost entirely by the mean, there are certain periods in which the fluctuations of each country's macroeconomic variables have a non-uniform incidence across the income distribution (see Figures (4) and (5) in Section (3)).

(2.) Macroeconomic variables such as GDP and consumption exhibit poor performance on predicting income movements, so other imputations of macroeconomic aggregates are necessary.

This paper proposes a method that combines the simplicity and standardizable features of the mean-scaling approach, with the distributional gains of more detailed simulations. Compared to the mean-scaling approach, our method is more complex to build but it offers the advantage of being able to determine distributional effects, a better forecast performance, and easiness to replicate across time and countries. First, we estimate a set of behavioral equations that forecast how individuals will change during the simulation period at the micro level for each country of analysis. These equations are used to predict the likelihood of increasing the number of schooling years, the changes in employment status, the expected change in labor income, and the likelihood of having formal work. Second, we use these results to simulate natural demographic changes to estimate how the target population will evolve during the forecasting period such in terms of their age, work experience, educational achievement, working structure, and earnings.[1] Third, we estimate the earnings of the simulated working population by imputing income to the transitioning working populations and adding marginal effects to those who remain in the same earning categories. This process results in a simulated distribution of the labor income that we supplement with assumptions of the non-labor income to yield a simulated distribution for the total income. Fourth, we use this data to estimate poverty rates at a country level.

Results suggest that, with the exception of Costa Rica, where non-labor income plays an important role among poor households, our model outperforms mean-scaling in predictive power. Moreover, if estimates are aggregated to estimate a regional poverty rate, the model's

[1] For working structure, we simulate how people move from working categories, and assume certain hierarchy in which movements are more likely to occur.

errors become remarkably small. As our method includes variations in unemployment (extensive margin of total labor income), it should be able to better capture changes in poverty under large employment shocks such as 2020. Although the lack of data does not allow to compare the models for all the countries, the results in Costa Rica and Dominican Republic for 2020 are in fact better in our model. In the case of El Salvador, our model overestimates by a bit more the change in poverty, for the same reason related to the extensive margin.

Given that our micro-simulation focuses on labor-income variables, other sources of income that affect poverty, are allowed to change independently with the rest of the micro-simulation. This makes our approach compatible with more detailed micro-simulations when needed. As our main objective includes a portable framework to nowcast poverty, we do not perform those kinds of variations, although the method does not rule them out.

The rest of this paper is organized as follows: Section (2) presents a brief summary of the key facts about poverty and income characteristics of Central American countries. Section (3) discusses the results and limitations of applying the mean-scaling approach in our countries. Because this is a method that closely relates to our main motivation, we use it as a benchmark to test our results. Section (4) describes the general characteristics of micro-simulations, and all the details of the micro-simulation we develop. Section (5) provides concrete results of applying our method, and a comparison with our benchmark, and finally Section (6) concludes.

2 Income poverty in Central America

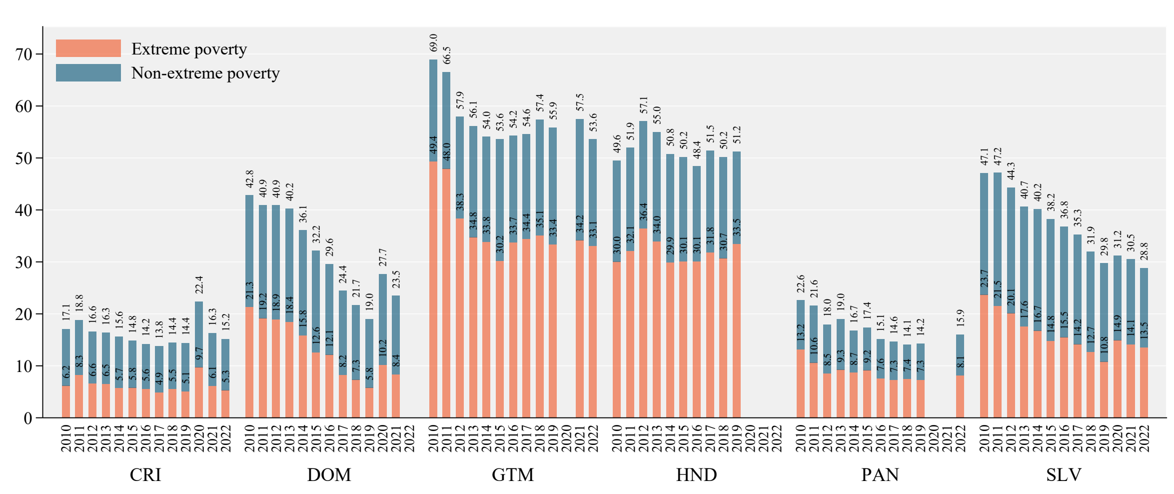
The economies of CAPDR exhibit a range of development levels. For instance, GDP per capita figures range from of 5,572 dollars[2] to 29,038 dollars[3]. This diversity in economic development has led to substantial differences in poverty rates across the region. Therefore, it is imperative to analyze social characteristics specific to each country to better understand poverty trends in the region.

Figure (1) captures the dynamic nature of poverty and extreme poverty across CAPDR, illustrating diverse trends and levels. Prior to 2020, the Dominican Republic and El Salvador demonstrated consistent reductions in poverty rates, in contrast to stagnation observed in other countries. However, the year 2020 marked a significant downturn, characterized by

[2]Honduras, constant 2017 PPP
[3]Panama, constant 2017 PPP

widespread income reductions, primarily driven by a steep decline in labor income across most countries. While increased public cash transfers partially cushioned this impact, total poverty rates surged notably in Costa Rica, the Dominican Republic, and El Salvador. The scenario shifted in 2021 with a rebound in labor income, playing a pivotal role in income recovery for low-income households. Concurrently, government transfers experienced a decline as emergency support measures, initiated during the pandemic's peak, began to phase out. Additionally, El Salvador, Guatemala, and Honduras consistently show the highest percentages of poverty and extreme poverty within the region.

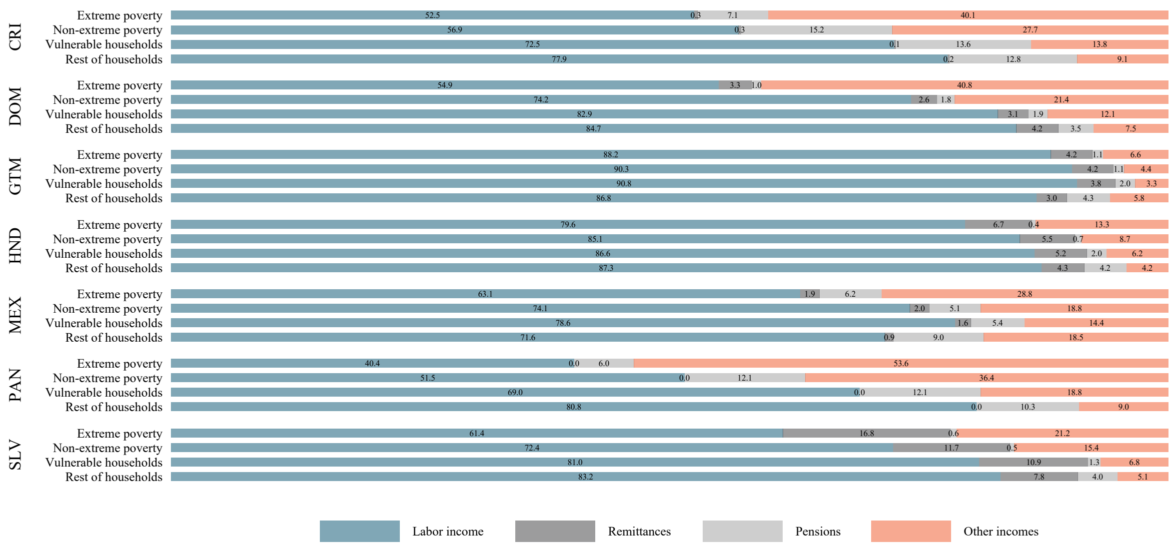
Figure 1: Evolution of poverty in Central America



Note: Households are classified by their daily income per capita (dollars in 2017 PPP). Extreme poverty: $[0.0 - 3.2)$; Non-extreme poverty: $[3.2, 5.5)$.

Also, a factor that makes countries different in terms of household incomes, is the composition of their incomes. Figure (2) provides a detailed depiction of household income composition across various income segments in CAPDR countries. It reveals that, for most of these countries, labor income constitutes the primary source of earnings. This observation underscores the critical role of labor income in poverty estimations, particularly in the context of significant employment fluctuations. However, non-labor income sources are significant for households in extreme poverty in Costa Rica and Panama. This suggests the potential impact of targeted government subsidies in these regions. Moreover, the figure highlights the significant contribution of remittance income in El Salvador, Guatemala, and Honduras, as well as in the Dominican Republic. In these countries, remittance income comprises a substantial portion of the total household income, ranging from 3.3% in the Dominican Republic to nearly 17% in El Salvador, for households experiencing extreme poverty. The non-labor income is a component of total income that can be a route to improve our method, that only relies on labor-income changes. Nevertheless, in the results section we show that, with the exception of Costa Rica, the modeling of labor-income seems sufficient to give good estimates.

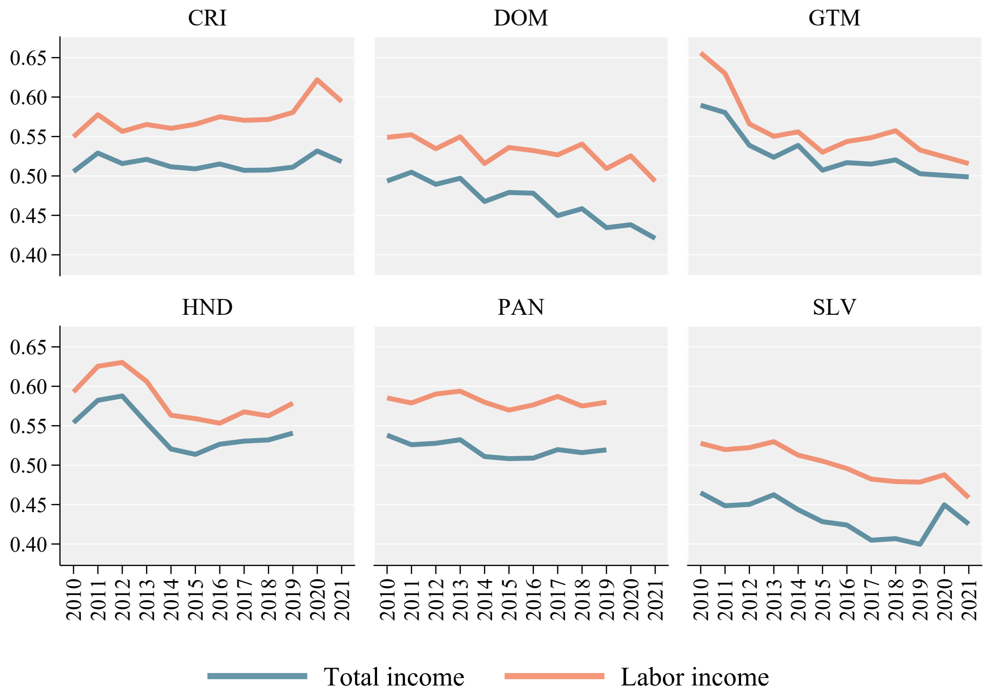
Figure 2: Households' income composition



Note: Households are classified by their daily income per capita (dollars in 2017 PPP). Extreme poverty: $[0.0 - 3.2)$; Non-extreme poverty: $[3.2, 5.5)$; Vulnerable households: $[5.5, 13.8)$; Rest of households: > 13.8.

Lastly, the income distribution has had important changes in the last decade, as shown in Figure (3). The Gini index of countries like Dominican Republic, Guatemala and El Salvador has a consistently falling trend, meaning that the reduction of poverty has been driven not only by changes in general income levels, but also by the distribution. As we discuss in the results section, this is a relevant fact explaining the improvement that our method offers with respect to alternatives that do not take into account distributional changes.

Figure 3: Evolution of Gini Index in Central America



3 A first benchmark: the *mean-scaling* method

One of our main motivations is to nowcast poverty in a way that is both reliable and easy to replicate between countries and different periods. For this purpose, a remarkable contribution is the method developed by Mahler et al. (2022), where just using GDP or other standard and accessible measures of welfare, they predict the change in the average income of surveys and scale all income accordingly[4]. After scaling the incomes in proportion to the change in the mean, poverty rates are estimated. Because of its advantages, and sharing a similar purpose with our work, this approach will serve as a benchmark to test the performance of our own model. Moreover, to have an even more strict benchmark, we will compare our model with a simple variation of the mean-scaling method that, using the same data, further improves its performance. That variation consists of estimating the change in the average income of different quantile groups of the distribution, and it is proven to be generally better in terms of poverty prediction.[5] The success of this method in predicting changes in poverty can be explained by the central role of the mean of the income distribution. They

[4]From now on, we will refer to this method as *mean-scaling*.

[5]A similar variation for Latin America is tested in Caruso et al. (2017), where they also find better results. In their paper, they call this variation *Quantile Growth Contribution* (QGC) method.

find that, in general, shifting other moments of the income distribution makes little difference in aggregate poverty predictions compared to only changing the mean. Authors show that a variable as simple as GDP per capita growth, can perform as good as models that use thousands of variables for prediction of the average income change.

The following results show up when the method is applied to CAPDR:[6]

(1.) It provides good estimates for some countries, but not all of them.

(2.) It is significantly improved by decomposing the distribution in quantiles.

(3.) Prediction errors between countries have low enough correlation, so it produces more precise predictions for the group's aggregate poverty, which is in line with the purpose of estimating global poverty the authors have in mind.

There are two main problems with the application of this method to the context of CAPDR.[7] First, although it might be generally true that the mean is enough for predicting changes in poverty, looking at countries individually, there are periods with rapid distributional changes induced by macroeconomic variations over time. As it was shown in Section (2), the countries of our analysis have shown significant decline in the inequality that is measured by the surveys. This distributional change can have large impact in poverty rates that are not accounted for by the mean. In Figures (4) and (5), it is shown that even in one period, a distributional change can have large impact relative to the impact that the mean alone has. This gets exacerbated when there are gaps in the years where surveys are available, as the distributional part of the poverty change gets more important as more periods ahead are being predicted.

[6] The set of countries where we apply this method are El Salvador, Costa Rica, Dominican Republic, and Honduras, because those are the countries that have enough annual series of household surveys. That is not required in principle to use the mean-scaling method, but is the case in which it works best, and thus serves as benchmark.

[7] It is likely that these problems would arise in many more countries, but we do not provide evidence of that in this paper, as we are just focused in CAPDR.

Figure 4: Decomposition of poverty changes from 2013

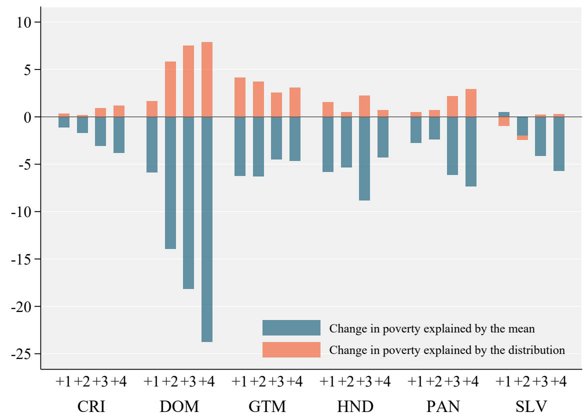
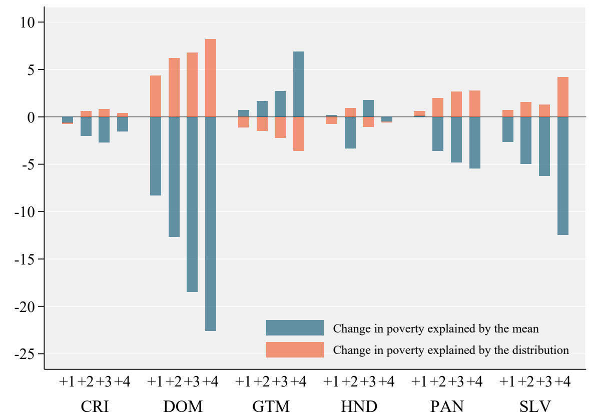


Figure 5: Decomposition of poverty changes from 2014



Note: The Figures show the decomposition of the poverty change one year, two years, three years and four years ahead from the base years, which are 2013 and 2014 respectively. The change in poverty explained by the mean corresponds to the mean-scaling estimated change using the observed change in the mean. The remaining part is the one explained by the distribution.

The first problem we have just mentioned has to do with the insufficiency of the mean to predict changes in poverty rates. The second problem is that, even if the mean was a sufficient predictor for changes in poverty rates, predicting the change in the mean of income is troubling in itself. The reason is that the link between macroeconomic data and the income reported by surveys is weak.

The inconsistency between income distribution in household surveys and macroeconomic data has been a topic of significant research interest in both advanced and developing economies, and poses a serious challenge to apply changes in GDP per capita to the income reported in surveys.[8] This raises concerns about the accuracy and reliability of income distribution data derived from household surveys even in countries with high statistical ca-

[8]Due to large discrepancies, between aggregate macroeconomic data and microeconomic sources, a growing body of literature has developed methods to infer the correct income distribution using additional data sources. This effort to address inconsistencies between incomes reported in surveys and aggregate data has become known as ”Distributional National Accounts” (DINA), and its used to create more reliable data on inequality across the world (see Alvaredo et al. (2016)). Following this approach for the United States and France respectively, Piketty et al. (2018) and Garbinti et al. (2018) have found substantial differences between the changes of income reported in household surveys and macroeconomic growth. As a consequence, both the level and the trends of inequality are not well captured by survey data. For a more general discussion, see Saez and Zucman (2020). Other relevant research conducted for high-income economies are Ederer et al. (2022) and Zwijnenburg et al. (2017). The first of these papers provides a standarized method to perform DINA analysis for countries without detailed tax data, and comes to similar conclusions than previous literature. The second one has contributed to this discussion by estimating gaps between microeconomic reported income and macroeconomic aggregates in OECD countries.

pacity.[9] CAPDR countries face similar challenges regarding the consistency between income reported in household surveys and GDP per capita. Over both short-term and long-term periods, there are noticeable discrepancies between these two measures, and the correlation between the real average income growth reported in surveys and the real GDP per capita growth is low across all countries in the region. For instance, Figures (6) and (7) show the case of Costa Rica[10], where real GDP per capita has very limited predictive power for average income growth in the survey. This discrepancy becomes more pronounced when focusing on the bottom half of the income distribution, which in turn holds greater importance in assessing poverty changes. Furthermore, these inconsistencies are not limited to total income versus GDP per capita; relevant components such as remittances also display substantial disparities with macroeconomic data, as highlighted in Appendix (A).

Figure 6: Changes in GDP and surveys' income (Costa Rica)

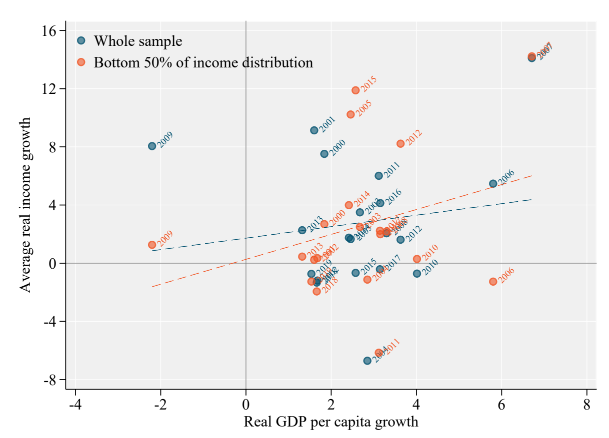
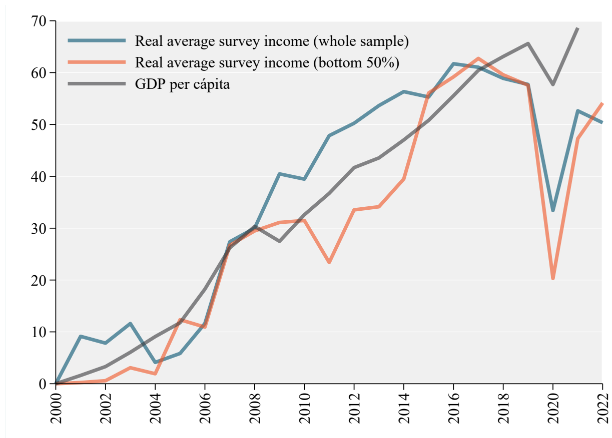


Figure 7: Level of GDP and surveys' income (Costa Rica)



Because of these two problems, the mean-scaling method has room for improvement if changes in the distribution are accounted, and more macroeconomic inputs that give additional relevant information are integrated to keep track of income. In our case, we introduce a more complex modeling of labor income, that responds to macroeconomic inputs.

[9]In the context of developing economies, the issue of inconsistent income distribution becomes even more pronounced. Burdín et al. (2020) have conducted research specifically focused on Latin America, revealing that household surveys in this region tend to substantially overestimate the decrease in income inequality. Furthermore, De Gregorio and Taboada (2022) have highlighted a significant underestimation of median income in Chile, with survey data falling short by 40%.

[10]For the rest of the countries see Appendix (A.2)

4 Our micro-simulation design

4.1 What is micro-simulation?

Simply put, a microsimulation refers to a set of modelling techniques that operate at the level of individual units (which could be persons, households, or firms, among others) in which we define a set of rules designed to simulate changes in behavior Figari et al. (2015). These results to assess distributional effects or calculate aggregated effects, introducing, if needed, interactions between individual units. This approach can be traced back to the 1950s, when Orcutt (1957) pioneered the use of these techniques to analyze the impact of social and economic policies. Since then, many applications emerged focusing within and outside the areas of policy evaluation which include tax-benefits Mirrlees and Adam (2010), labor markets (Blundell (2012); Keane (2011)), income distribution Stiglitz (2012), transportation Dowling et al. (2004), among others.

The contribution of these models is, mainly, to answer *what if...* type of questions based on a set of rules and shocks that can be calibrated to represent possible policy scenarios (for example, to determine how the poverty rates would change after a transfer program and its distributional effects). In this regard, microsimulation models can be classified into three (3) types based on how the rules affect individual behavior Harding (1996). The first type is static models, which define deterministic rules where the unit's characteristics remain constant over time. These are the most used models in the literature. The second type is dynamic models Li and O'Donoghue (2013), which *age* the units over time via introducing natural processes (such age changes) and are useful for modelling long-term policy effects such as the ones on pensions or health systems Hancock et al. (2013). The last type is behavioral models, which use micro econometric models of individual preferences to estimate the effects of policy changes on behavior.

Since Ortcutt's seminal work, micro-simulation techniques have been used for a variety of applications. However, given its ability to estimate distributional effects and behaviors at the individual level, the bulk of these models address policy-related questions. Table (3) in Appendix (B.1) presents a selection of applied micro-simulation articles and prioritizes listing papers with applications to the Latin American context. The topics range from tax reform analysis (early applications) to income distributional effects and poverty based on different policy shocks. One last relevant factor is that most, if not all, micro-simulation models are *ad-hoc* to the specific application or context. All applications cited part from a common principle (modelling behavior at the individual level) and create specific mathematical structures based on the research question and available data structure.

4.2 Overview of our method

Our methodological approach uses a two-step process that combines an array of statistical methods into four key components, transforming baseline raw data from household surveys and a series of macroeconomic indicators into predicted poverty rate estimates. It leverages the World Economic Outlook (WEO) projections of macroeconomic indicators of the real sector (International Monetary Fund (2023)) in tandem with time-series analysis models, machine learning algorithms, and econometric equations to forecast aggregate metrics of the labor market equilibrium. Then, it uses these forecasts to guide the distributional effects of the simulation process at the individual level.

To ensure that macro-micro consistency is maintained. The main components of the methodology are (i) forecasting methods that transform present macroeconomic and household survey datasets into projected data; (ii) a microsimulation framework for the labor market equilibrium that yields wages and labor income; (iii) a microsimulation framework for non-labor incomes; and, (iv) a translation mechanism to estimate poverty lines from macroeconomic variables.

For the first step, micro-simulation methods are used over data from household surveys to generate complete forward-looking simulated datasets with projected individual and household characteristics. In Parallel, time-series forecasting and imputation methods are applied to macroeconomic series to obtain projections for the key variables that drive income and poverty lines, mainly price levels and labor market metrics.

For the second step, the projected micro datasets and macro indicators are combined into an income generation framework that simulates the future income distribution at the individual levels, subject to the expected changes in aggregate macro indicators. To do so, a set of behavioral equations is used to estimate the probability of individuals belonging to employment categories and their respective expected earnings, given such categories. The forecasts of macroeconomic indicators of the labor market serve as bounding targets to constrain the aggregated population changes that may result from the simulation at the micro level, in terms of employment population and total income generated. From this, the distribution of labor income is generated. Lastly, non-labor incomes are simulated based on assumptions of the population and the underlying policies regarding pensions and government transfers, while poverty lines are projected using their conceptual definitions based on price index forecasts.

4.3 Forecasting methods for micro and macro conditions

The key goals of this process are to obtain a projected household survey dataset from the current data on individual and household characteristics, and forecasts for the macroeconomic indicators that are used to drive the expected changes in the income distribution and poverty lines. The resulting dataset of this process serves as the baseline population for simulating the future income distribution and the projected macroeconomic series are used as equilibrium parameters that guide such simulation.

For the household data, it is necessary to determine how individuals and households will look one period ahead into the future in terms of their demographic characteristics that impact their capacity to generate income, which includes household size and composition, age, gender, education, and other demographic characteristics that are time-variant. Alternatively, the macroeconomic forecast projects the variables that serve as bounds for the labor market simulation, which includes unemployment, labor participation rates, formal work, total wages, and inflation. Because the macro data is a collection of time series while the household dataset is a cross-sectional assessment of individuals and households in the present period, the projections are generated simultaneously and independently of each other.

4.3.1 Forecast of macroeconomic variables

As stated before, we perform a micro-simulation of the income distribution at the micro level but restrict that the aggregated results of that process should yield the forecast generated for the macro indicators. For instance, if the macro forecast for change in unemployment is 2, that means that after simulating the changes in the distribution of occupational categories of the labor market, the resulting projected unemployed population should be 2 percentage points higher than the baseline used for calculations. The key macro indicators we used to restrict the labor market simulation are change in unemployment, rate of change of labor force, rate of change in total labor income and rate of change in formal work.[11] Finally,

[11] To limit the complexity of the forecasts, and to allow for simplified analysis of scenarios, we used the growth in real GDP as a pivotal parameter. For that, we first estimate a baseline series of forecasts based on statistical methods and then we convert such forecasts into elasticities to GDP growth. In that way, growth in GDP becomes a guiding variable of the macro forecasts that could be changed to analyse multiple

poverty lines are defined based on changes in price indicators. Thus, a forecast for inflation would suffice to generate updated poverty lines.

Historical yearly data on GDP, inflation and unemployment are retrieved from the WEO (International Monetary Fund (2023)), whereas labor force participation, mean labor income and formal work rate are calculated directly from each household survey at the national level when available and filled with national aggregates from the International Labor Organization statistics ((International Labour Organization (2023a); International Labour Organization (2023b)). Unemployment data for Guatemala is not available in (International Monetary Fund (2023)), so it is retrieved from the World Bank Database (The World Bank, World Development Indicators (2023)). The baseline period of analysis for macroeconomic data is 1990 to 2022. Whenever there are gaps in the data, they are filled using linear interpolation while historical missing information is filled using the same method as for forecasting.

Additional data on consumption, investment, population growth, purchasing power factors and mean years of schooling is also gathered and used as exogenous inputs for the forecasting process. For the mean years of schooling, the variable is retrieved from UNDP (United Nations Development Program (2023)) where the data was available until 2021. Hence, a one-period-ahead forecast is also generated using the ARIMA time series modelling framework (Mills (2019)) to ensure that the baseline information is complete for the period of analysis.
For each indicator, a different forecasting method is implemented depending on data availability, theoretical considerations and back-testing performance of statistical methods evaluated. The following summarizes the forecasting process:

- **Exogenous variables, GDP growth, and inflation**: forecasts of these variables are taken directly from (International Monetary Fund (2023)) which uses their in-house methodology to provide projections until 2028. As these are pivotal indicators in the methodological process, they are taken as they are without modifying, calibrating or overriding any of the values from the original series.
- **Change in labor force**: the output series to forecast for this indicator is the percentage change in total workers of each year-country combination. While the original data comes from household surveys, whenever missing, it was replaced with data (International Labour Organization (2023a)). If there were still gaps remaining, these were

macroeconomic scenarios in a summarized manner.

filled through linear interpolation methods. For the final forecasts, a machine learning algorithm based on the random forest model (Bonaccorso (2018)) is used to predict the percentage of change in the labor force. All macroeconomic variables available, as well as the forecasts for GDP and inflation, are used in the modelling process.

- **Change in unemployment**: the output series to forecast is the percentage of change in the unemployment rate. Original data is generated from household surveys but gaps are replaced with (International Labour Organization (2023a)) if available, or linear interpolation. For the final forecasts, a machine learning algorithm based on the random forest model (Bonaccorso (2018)) is used to predict the percentage of change in the unemployment rates. All macroeconomic available variables are used in the modelling process, as well as the forecasts for the change in labor force, GDP and inflation.

- **Change in labor income**: for these forecasts, the growth in the mean labor income is used as the baseline data series from household surveys. Gaps are replaced with direct estimates of the growth rates from (International Labour Organization (2023b)) when available, and linear interpolation if not. The forward-looking forecasts are generated using a machine learning algorithm based on an ensemble of models, used to predict the percentage of change in the unemployment rates. All macroeconomic available variables are used in the modelling process, as well as the forecasts for the change in labor force, change in unemployment, GDP and inflation.

- **Change in formal work**: for these forecasts, the percentage of change in the total number of formal workers is the target variable, which is generated from the survey data by default, but is possible to use country-specific datasets as the source in custom versions of the model. When using survey data, the forecasts are generated using a machine learning algorithm based on an ensemble of models. All macroeconomic available variables are used in the modelling process, as well as the forecasts for the change in labor force, change in unemployment, change in labor income, GDP and inflation.

  Once estimated, the specific model for each equation is used to generate its forecast but since the goal is to pivot all the labor market variables to the expected GDP growth, the elasticities of each indicator for a given country i are calculated using the expression:

$$\eta_{i,t}^{\mathrm{GDP},x} = \frac{\Delta\% x_{i,t}}{\Delta\%\mathrm{GDP}_{i,t}}$$

Where: $\eta_{i,t}^{\text{GDP},x}$ is the expected percentage change in indicator x given a change of 1% in real GDP in year t for country i. These elasticities can be used for performing complete forecasts and simulations for different scenarios of economic growth, using real GDP as the pivotal parameter. However, in a practical sense, they are not used in our key calculations. The macroeconomic indicators serve two key functions in the methodology. First, the variables related to the income distribution (GDP) and labor market (unemployment, labor force participation, mean labor income and informal rate) are used as parameters that guide both income microsimulation processes by setting aggregate boundaries to the expected outcomes from the behavioural equations. Secondly, because they are also defined in their elasticity form, they would allow for the analysis of simulation scenarios based on a simple parameter, the percentage change in real GDP.

4.3.2 Simulation of individual and household characteristics

For this component, we need to generate an updated survey for the forecasting period. That means, advancing the population in the current survey for the number of years that the simulation process is intended to model. However, since the key goal of the new survey data is to simulate new income distribution, not all characteristics in the survey are relevant for the analysis. Also, some characteristics are time-invariant, which would not require any method to be projected. Our approach simulates the following characteristics:

- **Age**: the population is aged by the total period of the forecast. For example, if the projection is one year, then all individuals in the population are aged by one additional year. No mortality rate is assumed, so the total number of individuals in the sample does not change.

- **Years of schooling**: The expected number of years of schooling is calculated on each period of the forecast for all individuals currently studying, based on an ordinal logistic regression equation. Those who obtain a higher number of years than the reported one receive one additional year at each forecasting period, while those with equal or lower years remain as they are. Individuals not currently studying remain with the same number of years, irrespective of their age. That means that we do not account for new entrants to the population in schooling age. However, since the key goal of this simulation is to project the dynamics of the labor market, there is no impact on the key model outcomes as new entrants of the schooling population are not part of the labor market population in a formal sense.

- **Expected years of experience**: we use the traditional approach that defines the expected years of experience as the age minus the years of schooling minus six, with a lower limit of zero (Mincer (1975)). Given that the updated expected years of experience are estimated using the same definition, the projected expected years of experience are generated using the updated age and years of schooling.

- **Demographic characteristics**: gender, marital status and other demographic status are assumed to remain the same for all the population. Additionally, birth and mortality rates are assumed to be equal and symmetrical, meaning that no changes are made to the underlying age structure or size of the population except changing the age for the time of the forecasting horizon.

- **Household characteristics**: household composition is assumed to remain constant and no changes are made to the size or members of the household. That includes assuming that there is no migration to the household or outside of them, as well as not allowing for the formation of new households.

Once each step is completed the resulting output is a simulated dataset that has advanced the household survey population into the future for the number of periods required given the forecasting horizon. This newly simulated data becomes the baseline on which the simulation of the income distribution takes place.

4.4 Simulation framework for labor income

As noted above, the distinction of labor and non-labor income in the simulation process allows for more precise and simplified frameworks since the mechanism generating each of them is conceptually different. In the case of labor income, it can be seen as the aggregate of all the individual incomes generated by the individuals that participate in the labor market, in any form, and whose earnings can be expressed as a function of their characteristics. Hence, if the newly updated characteristics of individuals resulting from the simulation process applied to the household survey are used as input for such function, it will yield simulated labor earnings for the period of analysis.

The main challenges for this process are to generate such a function that could map simulated individual characteristics into simulated earnings, and to maintain the consistency of the resulting distributions of income and working population with the macroeconomic forecasts. Due to the structure of the labor market, our approach estimates this function in a compound manner. First, we determine the employment category of each person in the population, based on their simulated characteristics and their previous employment status and earnings. This is because the earnings observed from data depend on each individual's decision to participate in the labor market, as well as the market itself allowing them to. That is, individuals can be employed if they decide to participate in the market and succeed; unemployed, if they are willing to participate but are unsuccessful in landing a job; or inactive, if they decide not to participate at all. Additionally, individuals may be able to participate in the labor market, but not in the way and intensity that they are willing to, pushing them into informal work.

Even though we observe these decisions in the household data, in the simulated period we cannot determine what will be the choice of each individual, given their new characteristics. Mainly, because we do not have enough data to account for it, as this can change not just by demographic and human capital characteristics, but also because of unobserved individual preferences; and the limitation of repeated cross-section data not able to capture individual dynamics. Hence our first step is to determine the new structure of the working population given the changes induced to the overall population by simulating their progression into the forecasting horizon. To do this, we use a set of behavioral equations regarding the probability of individuals belonging to each employment type, and an algorithm to assign individuals to each category. Then, we estimate the new earnings of this updated population based on another set of behavioural equations for earnings that are conditional on the simulated employment categories.

4.4.1 Simulated labor populations

The first element that determines changes in the labor income distribution is the transition in the structure of the labor force. Initially, we identified that there would be new entrants to the labor force due to the natural transition in age and the productivity life cycle of individuals. That is, new individuals enter the age where they can be part of the labor force and may decide to participate in employment activities. From a purely practical sense, there is no specific discrete point in time where individuals enter the labor force, although there are specific ages where the entrance is higher and more likely. However, due to the nature of household surveys, the legal age of work is a key turning point where individual's data regarding working activities is gathered. Hence, when advancing the age of the population in the simulation process, the employment characteristics of those individuals that move from non-working to working age are also simulated because they would be observed in the simulated period. To do so, we estimate the marginal probabilities of each employment category for those individuals turning into the age of working in the simulated dataset.

We use the definitions for working age of 15 years and older, and categories of employment structure consisting of employed formal, employed informal[12], unemployed and inactive, following (International Labour Organization (2023a) In that sense, we estimate the probabilities of each 14-year-old in the base dataset (thus turning 15 in the immediate simulated period) for each employment category and then assign them to the one with the highest probability. These probabilities are calculated by inputting the simulated characteristics of the new entrants into a set of country-specific multinomial logistic regressions that are estimated using the historical repeated cross-section data on 15-year-old individuals. The equations are calibrated using only data from the population of this age group, since it is deemed as having idiosyncratic characteristics with respect to the general working force. A separate equation is used for each level of employment category: (i) formal and informal, given individuals are employed; (ii) employed and unemployed, given individuals are active; and, (iii) active and inactive, given that individuals are of working age. It is worth noting that the probabilities used are marginal since we are not using panel data. Therefore, it is likely that we overestimate the number of individuals moving to the employed or unemployed categories, since they are new disruptive status (for example, for individuals who are currently studying), while the inactive category would be a preservation of the status quo. This is a limitation due to the nature of the data available. We attempt to reduce the impact by introducing key variables on the individuals and their household characteristics that would help to identify possible shifts in the employment category. Also, due to the small size of this age group, we deemed the impact of this limitation as minimal.

The following step is determining the employment categories of individuals that are already of working age (15 years and older in the baseline data). For this, we follow a similar process to the one used for new entrants by estimating the probabilities of every individual belonging to each employment category and then assigning them to the one with the highest likelihood. However, there are two important caveats that we implement to ensure that the overall transition is consistent at the micro and macro levels.

On the micro side, because this population is already of working age, their true probabilities in the simulated period are not marginal but conditional on the category that they already belong in the baseline data. And given the long-term nature of the labor market, transitions between employment categories are less likely to occur. For example, someone

[12] We used ILO data for all countries modelled within the micro-simulation structure which are Costa Rica, the Dominican Republic, El Salvador, Honduras, and Panama

who is employed would be, on average, more likely to remain employed than become inactive or unemployed, given that there are no changes in their characteristics. Similarly, a retired individual is more likely to remain retired than unemployed. Unfortunately, as stated above, our data available is a set of repeated cross-sections that does not observe individual transitions and forms in which any marginal probability estimate would be higher than the actual probability of each category. Therefore, using the same approach that we did for the new entrants would result in more transitions than actually would happen in the real market.

Regarding the macro side, as noted in (4.3.1), the macroeconomic forecasts of the aggregated equations are used as guidelines for the microsimulation to ensure consistency in general equilibrium terms. That means that if the forecasts state that the unemployment rate will increase by 2%, the total population of unemployed individuals in the microdata set needs to amount to that percentage of the active working force. If the employment categories assigned are based solely on probabilities, the resulting population will not necessarily match our projected metrics of the labor market, since it will depend only on the distribution of the estimated probabilities from the microdata.

Given the above, we first estimate the marginal probabilities using the same approach as we did for new entrants, by inputting individual characteristics into behavioural econometric equations based on the multinomial regression model. Then, we calculate the individual probabilities of each employment category, but instead of assigning groups using these probabilities, we assign category-specific rankings to the individuals based on the likelihood of each category. That way, we generate 4 rankings, one for each employment category, that identify which individual is more likely to belong to the said group. We later use these rankings to assign individuals to employment groups, up to the point until we reach our macroeconomic target. For example, if we predict that formal employment will increase to 100,000 workers, then we assign the top 100,000 individuals with the highest probability of being formally employed to that group. However, an important consideration is that our macroeconomic metrics have set targets for several populations (labor force, formal work, and unemployment). Thus, our process needs to ensure consistency of all categories, as well as simultaneously assign individuals to mutually excluded groups.

Theoretical assessments of the labor market indicate that some transitions can be asserted as more likely than others, given a macroeconomic shock. We implement a sequential algorithm that assigns individuals to employment categories based on our assessment of which movements are more likely to occur, but also minimizing the required income imputations in the simulated groups. For that, we establish a hierarchy of group assignations and transitioning patterns, depending on the guiding macroeconomic forecasts. The following figure describes our algorithm.

Under this approach, our assignation sequences are as follows:

- If the expected change in the labor force is positive:

  We expect individuals to enter the labor force, so we transition the inactive people with the highest probability of being unemployed into the unemployment category until the labor force matches the macroeconomic forecast. Now we move the individuals with the highest probability of being informally unemployed into the informal work category until we reach our unemployment rate target from the macroeconomic forecast. Finally, we push individuals from informal to formal work, based on their probability of being formal, until we meet the macroeconomic forecast for formal work rate.

- If the expected change in labor force is negative:

  We expect individuals to exit labor force, so we first move individuals between formal and informal employment, until the formal work rate reaches the macroeconomic forecast, based on their rankings from the probability of being in formal employment. Now, from the new pull of informal workers, we transition individuals into unemployed, until the target for the employed population is met (given by the unemployment rate forecast), using their rankings based on the probabilities of being unemployed.

This approach has a key limitation in assuming which transition movements take place in which order, which limits the evolution of the working population in a fixed manner, given the expected change in the labor force. However, this would simplify the shifts of individuals across categories, when there is no available data to estimate true transition rates. As mentioned earlier, our goal is also to reduce the number of imputations made on labor income, which are part of the calculations of updated income described below.

4.4.2 Updated labor income

Once the baseline working population has been simulated and transitioned one period ahead in the future, we would require to update their labor income based on their new categories of employment and their updated individual characteristics that influence their income generation capabilities, which include their education, new work experience, age, the new distribution of household employment, etc. However, in our updated working groups, some transitions have a structural impact on the distribution of labor income. The following table summarizes the labor income assignment made based on such transitions.

Table 1: Transitions between employment status and imputations

	Employed	**Unemployed**
Employed	Baseline income + marginal effects simulated characteristics	Income set to 0
Unemployed	Imputed income based on simulated characteristics	Income preserved as baseline

Following this, the labor income of employed workers who retain their status is estimated using their value in the baseline period but adding the marginal effect due to the changes in their updated characteristics. For unemployed and inactive workers who join the workforce, their total income is imputed, as they do not have any value in the baseline period. For the employed group that is reclassified as unemployed or inactive, their income is automatically set to zero. Whereas for the unemployed and inactive who keep their category, their income from the baseline period is maintained. This is because in some rare cases, we found individuals in these groups to have a small amount of registered labor income. Even though this could be deemed as a measurement error from the surveys, we aim to preserve the official income distribution as much as possible, thus, the incomes are preserved as they are.

The marginal effects and the imputed income are calculated through a set of country-specific econometric equations that predict labor income given individual characteristics. These equations are estimated from the historical data. In the case of the marginal effects, we predict the labor income using the simulated and baseline characteristics of individuals and we calculate the difference between them. Then, this difference is added to the baseline income. For the imputed income, the complete set of simulated characteristics is input in the equation and the resulting value is set as the imputed income. Repeating this process for all individuals in each group yields the simulated income distribution in the updated dataset.

4.5 Simulation framework for non-labor income

In our analysis, non-labor income encompasses government subsidies, pensions, remittances and other transfers. Because of the numerous exogenous factors that impact the generating function of non-labor income, and since most of its components are specific to the legislation and macroeconomic environment of each country of analysis, an overarching framework for simulating it requires a broad set of assumptions on household composition and government policies. Moreover, we noted in the previous section that forecasting remittances from individual microdata is a methodological challenge that usually leads to contradictions when contrasted with national accounts data (see Figure (13) in Appendix (A.3)). Due to these challenges, and given the low relevance noted from this type of income over the total household incomes for most of the population, we decided to simplify our approach in a baseline scenario analysis by simply scaling the non-labor income of each individual by the inflation rate for the forecasting period. Nevertheless, our model has been designed to avail for the use of assumptions for each type of non-labor income in several forms if an analysis of scenarios is intended.

The final income we estimate is the sum of the simulated labor income from the previous framework, and the scaled non-labor income, for each individual. The aggregation of these incomes results in the simulated income distribution for the forecasting period.

4.6 Projected poverty lines

The final element of the simulation is the projection of poverty lines. By definition, poverty lines are estimated using price and purchasing power parity indexes to come up with a homogeneous standardized line for all countries.[13] Given that, we use the inflation rate of each country as a direct estimate of the change in the poverty line. That is, the predicted poverty line $\hat{\text{PL}}$ is estimated using the expression: $\hat{\text{PL}}_{i,t+1} = \text{PL}_{i,t}(1 + \hat{\pi}_{i,t+1})$, where $\hat{\pi}_{i,t+1}$ is the predicted inflation rate for country i in the immediate next year. Once projected, poverty lines are used directly with the projected income distribution from simulating labor and non-labor income.

5 Results

The model is tested in Costa Rica, Dominican Republic, El Salvador, Honduras and Panama. The equations and processing components of the model are estimated using data from 2000 to 2021, and its performance is compared with the best possible variation of the Mahler et al. (2022) model in each country. In Figure (9), the observed changes in poverty rates are shown over time, and compared to the predicted values. In general, the predictions exhibit a good fit to the data. Also, the correlation of the errors between countries is sufficiently

[13] The poverty rate we use for our results is the percentage of households with daily income per capita (dollars in 2017 PPP) below 5.5.

low to get a precise estimate of the group's aggregate poverty rate. As can be seen in the figure, the predictions are close to the observed value with only a few exceptions. Table (4) in Appendix (B.2) shows the whole series of observed and estimated changes from 2000 to 2020.[14]

Figure 8: Observed and estimated changes in poverty rates

Comparing our method with our benchmark, we find that, with the exception of Costa Rica, where non-labor income plays an important role among poor households, our model outperforms mean-scaling in predictive power, for the 2000-2019 period.[15] The comparison is shown explicitly in Table (2), in which each column contains the estimation error $\hat{\varepsilon}_t = y_t - \hat{y}_t$, where y_t and $\hat{y}_t$ are the observed and predicted poverty rates respectively. Columns are labeled *Mic.* for the errors of our method, and *MS* for the mean-scaling method.[16] The performance of each option is summarized with the absolute mean deviation[17]. The measure of performance is better in all countries except for Costa Rica, and it leads to particular improvements in the cases of El Salvador and Dominican Republic. This is an interesting

[14]Due to the lack of a long enough series of surveys, both our method and the benchmark method have bad performance in predicting poverty rates of Guatemala and Nicaragua (see Table (5) in Appendix (B) for specific information about surveys availability). The results shown are for the rest of CAPDR.

[15]This result is robust to the inclusion of 2020 in the countries where household surveys were available, but we present the result from 2000-2019 to avoid any potential distortions introduced by the pandemic.

[16]The best variation of the mean-scaling method is used as benchmark. With the exception of Costa Rica, the best variation is dividing the households in quintiles, and estimating the change in income of each quintile using an OLS regression over the macroeconomic series of consumption per capita. For Costa Rica, the simple method, without quintiles division works better.

[17]The formula for the absolute mean deviation is $\frac{1}{T}\sum_{t=1}^{T} |\hat{\varepsilon}_t|$.

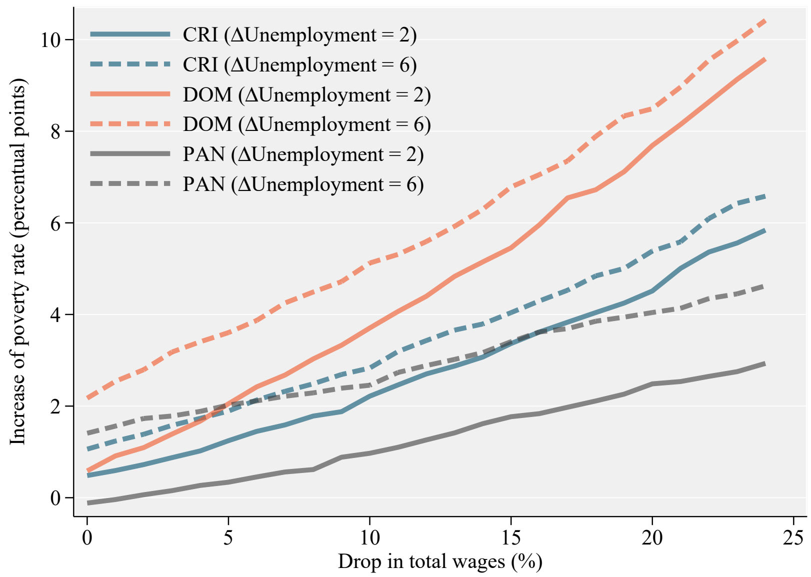
fact because those two countries are the ones with the greatest reduction of labor-income inequality, which is better captured by the simulation.

Table 2: Comparison between our micro-simulation and the mean-scaling method

Year	**CRI**		**DOM**		**HND**		**PAN**		**SLV**	
	Mic.	MS	Mic.	MS	Mic.	MS	Mic.	MS	Mic.	MS
2000	2.88	1.41	.	.	.	.	-0.09	-1.13	0.99	0.71
2001	-7.62	-0.50	-0.68	-1.57	.	.	-4.31	-5.75	-0.77	-0.96
2002	-0.15	-1.96	-0.27	1.33	0.60	-0.72	1.92	1.34	-0.60	-0.15
2003	3.83	0.23	0.85	-0.19	1.16	1.95	0.11	0.64	0.40	1.13
2004	0.65	-1.92	4.98	-8.15	-1.84	-2.55	-2.54	-1.00	-0.04	-3.37
2005	6.13	1.81	-0.67	0.28	0.17	0.97	0.14	-0.72	-3.35	-4.78
2006	-2.24	-0.55	-0.94	-1.53	2.36	2.46	0.26	-0.53	2.86	3.47
2007	3.43	3.56	-2.55	-0.94	0.84	2.83	1.39	2.29	1.85	1.05
2008	-2.89	0.62	-2.15	-2.09	1.23	2.26	1.17	0.58	-2.10	-3.99
2009	3.52	-0.37	-0.42	-0.17	-0.85	1.33	-0.50	0.21	-0.09	0.59
2010	4.17	-0.67	-2.08	-2.79	-4.74	-3.25	0.09	1.01	-0.58	-0.11
2011	-3.54	-2.69	0.27	2.02	-0.46	-2.46	-0.13	0.29	-0.04	-0.94
2012	0.97	1.00	0.07	1.20	-5.00	-5.87	2.67	2.72	1.74	1.97
2013	-0.17	-0.37	-1.17	2.01	1.05	1.62	-1.72	-1.97	2.35	2.63
2014	-0.43	0.09	0.35	2.36	4.24	4.44	0.80	1.28	0.67	-0.04
2015	-0.02	0.13	0.31	0.86	0.32	0.24	-2.02	-1.32	-1.42	0.33
2016	1.14	0.16	-0.26	0.46	-0.01	1.09	0.50	1.12	0.13	0.22
2017	0.31	-0.05	1.75	4.70	-4.28	-4.00	-0.61	-0.05	0.03	0.51
2018	-1.51	-1.43	0.27	0.73	0.36	-0.10	-0.24	-0.03	1.10	2.23
2019	-0.54	-0.47	0.74	1.54	-1.51	-1.70	-1.39	-0.97	0.06	1.08
2020	-4.22	-8.36	-4.27	4.56	.	.	.	.	0.51	0.44
	Absolute mean deviation (2000-2019):									
	2.31	1.00	1.09	1.84	1.72	2.21	1.13	1.25	1.06	1.51

Also, due to the simulation of the extensive margin, our model should better capture changes in poverty under large employment shocks such as 2020. Unfortunately, the lack of data does not allow to compare the models for all the countries. However, in Costa Rica and Dominican Republic, the 2020 prediction is in fact better in our model, and in the case of El Salvador, it overestimates by a bit more the change in poverty, for the same reason related to the extensive margin. This is not systematic evidence, but is suggestive of a better modeling of distributional determinants of changes in poverty, that makes our method particularly useful for periods of relevant stress in the labor markets. To make this point clearer, Figure (9) shows how the predicted increase of poverty rates changes when, for a fixed drop in total wages (i.e. the sum of all wages), a different increase in unemployment is used as input. This exercise shows that a same drop in total wages can have a much larger impact if the drop is driven by the extensive margin (higher unemployment). This creates a non-linearity that we argue is better to capture the empirical patterns of distributions and poverty.

Figure 9: Simulated poverty changes under different combinations of intensive-extensive margins



Finally, although the results are satisfactory, the fact that some countries have large components of income that do not come from labor, gives room for improvement in the performance of the micro-simulation. However, in the implementation of the micro-simulation, the treatment of other sources of income, which consisted only in adjusting for inflation, is completely independent from the labor-income simulation, which makes it possible to add another *layer* of simulation, if enough inputs for the other income are available in a given country or period. For example, in López et al. (2020), the estimations for Central America included fiscal transfers that were used as relief during the crisis. Those kinds of exercises can be perfectly done on top of this micro-simulations. In other words, applying our method does not rule out any other changes in incomes that do not provide from income. However, our purpose was to have a portable method, which is why our design has not included those other possibilities.

6 Conclusion

Knowing the poverty rates of countries depends on the availability and usually long processing times of household surveys data. To have a reliable estimate of those poverty rates before they are available, micro-simulations are a useful tool. However, these methods are generally hard to implement with flexibility and having an estimate for a large group of countries and for each period without great efforts is a significant challenge. Because of that, some alternatives make the estimations easier to standardize across countries and periods, with a significant lost of accuracy.

We propose a method that micro-simulates changes in labor-income using macroeconomic inputs that are widely available and standard, and can be replicated in a practically automated manner, while significantly improving the performance of simpler alternatives. This method affects the labor status of individual units of household surveys, according to macroeconomic data that has to fit the changes, and demographic transitions that are used as inputs. Concretely, our results give better performance for all the countries in which we test the method, with the exception of Costa Rica, where the presence of other sources of income make a micro-simulation that is purely based on labor income insufficient. However, our method is fully compatible with simulations of other sources of income that can be made on top of the base simulation for the labor market. Finally, our method captures the effect of unemployment on poverty, which amplifies the effect of total wages. This should allow the micro-simulation to outperform methods that do not include changes in unemployment, for periods of large stress in the labor market, as the 2020 pandemic.

References

Absalón, C. and Urzúa, C. M. (2012). Distributive effects of the 2010 tax reform in mexico. *Fiscal Inclusive Development: Microsimulation Models for Latin America*, page 101.

Acevedo, I., Castellani, F., Flores, I., Lotti, G., and Székely, M. (2020). Social effect of covid-19: Estimates and alternatives for latin america and the caribbean. *Latin American Economic Review*, 29(1):1–44.

Alvaredo, F., Atkinson, A., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2016). Distributional national accounts (dina) guidelines: Concepts and methods used in wid. world.

Amarante, V., Bucheli, M., Olivieri, C., and Perazzo, I. (2011). Distributive impacts of alternative tax structures: the case of uruguay. *Documento de Trabajo/FCS-DE; 9/11.*

Arancibia Romero, C., Dondo, M., Jara, H. X., Macas Romero, D., Oliva, N., Riella, R., Rodriguez, D., and Urraburu, J. (2019). Income redistribution in latin america: A microsimulation approach. Technical report, WIDER Working Paper.

Bargain, O. and Callan, T. (2010). Analysing the effects of tax-benefit reforms on income distribution: a decomposition approach. *The Journal of Economic Inequality*, 8:1–21.

Blundell, R. (2012). Tax policy reform: The role of empirical evidence. *Journal of the European Economic Association*, 10(1):43–77.

Bonaccorso, G. (2018). *Machine Learning Algorithms: Popular algorithms for data science and machine learning*. Packt Publishing Ltd.

Brum, M. and De Rosa, M. (2021). Too little but not too late: nowcasting poverty and cash transfers' incidence during covid-19's crisis. *World Development*, 140:105227.

Burdín, G., Vigorito, A., de Rosa, M., and Vilá, J. (2020). Was falling inequality in all latin american countries a data-driven illusion? income distribution and mobility patterns in uruguay 2009-2016.

Cabezas, G. and Acero, C. (2011). Effects of the ethical family income on labor participation, income distribution and poverty. *Fiscal Inclusive Development: Microsimulation Models for Latin America*, page 71.

Caruso, G. D., Lucchetti, L., Malasquez, E. A., Scot, T., and Castaneda, R. (2017). But? what is the poverty rate today? testing poverty nowcasting methods in latin america and the caribbean. *What is the Poverty Rate Today*.

Castañón-Herrera, A. and Romero, W. (2012). A microsimulation model for guatemala: the case of direct and indirect taxes. *Fiscal Inclusive Development: Microsimulation Models for Latin America*, page 87.

De Gregorio, J. and Taboada, M. (2022). *Median Labor Income in Chile Revised: Insights from Distributional National Accounts*. Universidad de Chile, Departamento de Economía.

Dowling, R., Skabardonis, A., Alexiadis, V., et al. (2004). Traffic analysis toolbox, volume iii: Guidelines for applying traffic microsimulation modeling software. Technical report, United States. Federal Highway Administration. Office of Operations.

Ederer, S., Ćetković, P., Humer, S., Jestl, S., and List, E. (2022). Distributional national accounts (dina) with household survey data: Methodology and results for european countries. *Review of Income and Wealth*, 68(3):667–688.

European Comission (2024). Euromod - tax-benefit microsimulation model for the european union.

Figari, F., Paulus, A., and Sutherland, H. (2015). Microsimulation and policy analysis. In *Handbook of income distribution*, volume 2, pages 2141–2221. Elsevier.

Garbinti, B., Goupille-Lebret, J., and Piketty, T. (2018). Income inequality in france, 1900–2014: evidence from distributional national accounts (dina). *Journal of Public Economics*, 162:63–77.

Hancock, R., Malley, J., Wittenberg, R., Morciano, M., Pickard, L., King, D., and Comas-Herrera, A. (2013). The role of care home fees in the public costs and distributional effects of potential reforms to care home funding for older people in england. *Health Economics, Policy and Law*, 8(1):47–73.

Harding, A. (1996). *Microsimulation and Public Policy: Selected Papers from the IARIW Special Conference on Microsimulation and Public Policy, Canberra, 5-9 December, 1993.* Elsevier.

International Labour Organization (2023a). ILO modelled estimates database, ILOSTAT [Labor Force Statistics].

International Labour Organization (2023b). ILO modelled estimates database, ILOSTAT [Wages and Working Time Statistics].

International Monetary Fund (2023). World Economic Outlook Database [Data set].

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, 49(4):961–1075.

Larrañaga, O., Encina, J., and Cabezas, G. (2012). A microsimulation model of distribution for chile. *Fiscal Inclusive Development: Microsimulation Models for Latin America*, pages 51–70.

Li, J. and O'Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of microsimulation*, 6(2):3–55.

López, A., Eggers, C., and Barrios, J. (2020). Poverty and inequality in the face of the pandemic. In *Inequality and Social Discontent: How to Address Them through Public Policy: Economic Report on Central America, the Dominican Republic, Haiti, Mexico, and Panama*, pages 6–20. Inter-American Development Bank.

Mahler, D. G., Castañeda Aguilar, R. A., and Newhouse, D. (2022). Nowcasting global poverty. *The World Bank Economic Review*, 36(4):835–856.

Mills, T. C. (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. Academic press.

Mincer, J. (1975). Education, experience, and the distribution of earnings and employment: an overview. *Education, income, and human behavior*, pages 71–94.

Mirrlees, J. A. and Adam, S. (2010). *Dimensions of tax design: the Mirrlees review*. Oxford University Press.

Molina, G. G., Montoya-Aguirre, M., and Ortiz-Juarez, E. (2022). *Addressing the cost-of-living crisis in developing countries: Poverty and vulnerability projections and policy responses*. United Nations Development Programme.

Molina, G. G., Ortiz-Juarez, E., and NETWORK, U. G. P. (2020). *Temporary basic income: Protecting poor and vulnerable people in developing countries*. United Nations Development Programme.

Navicke, J., Rastrigina, O., and Sutherland, H. (2014). Nowcasting indicators of poverty risk in the european union: A microsimulation approach. *Social Indicators Research*, 119:101–119.

Nogueira, J. R. B., Bezerra, R., and de Souza, E. S. (2011). A brazilian tax-benefit microsimulation model. *Microsimulation Models for Latin America*, pages 19–50.

Orcutt, G. H. (1957). A new type of socio-economic system. *The review of economics and statistics*, 39(2):116–123.

Piketty, T., Saez, E., and Zucman, G. (2018). Distributional national accounts: methods and estimates for the united states. *The Quarterly Journal of Economics*, 133(2):553–609.

Rueda, S. (2021). Microsimulaciones para anticipar decisiones en pobreza. Technical report, PNUD.

Saez, E. and Zucman, G. (2020). The rise of income and wealth inequality in america: Evidence from distributional macroeconomic accounts. *Journal of Economic Perspectives*, 34(4):3–26.

Stiglitz, J. E. (2012). *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

Tejerina, L. and Muñoz, L. (2015). 20 años de reducción de pobreza y desigualdad en el salvador. Technical report, Inter-American Development Bank.

The World Bank, World Development Indicators (2023). Unemployment, Total (% of Total Labor Force) (Modeled ILO Estimate) - Guatemala.

United Nations Development Program (2023). Human Development Reports, Data Downloads.

Zegarra, E. and Tuesta, J. (2009). Shock de precios y vulnerabilidad alimentaria de los hogares peruanos. MISC.

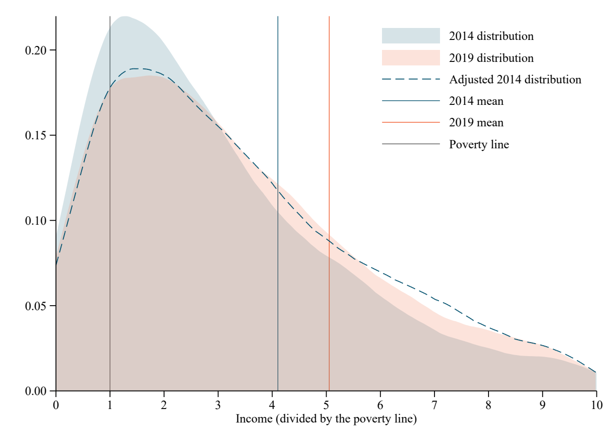
Zwijnenburg, J., Bournot, S., and Giovannelli, F. (2017). Expert group on disparities in a national accounts framework: Results from the 2015 exercise.

A Additional figures

A.1

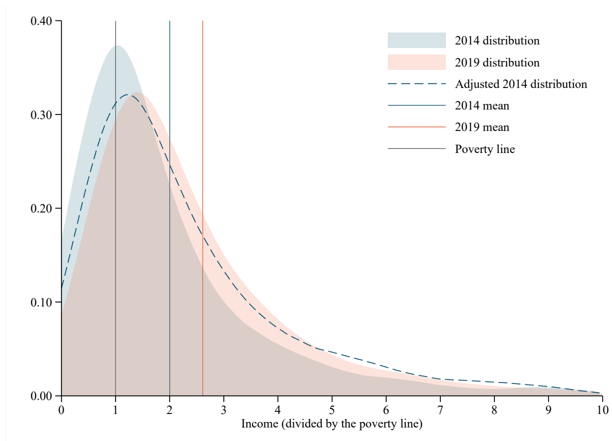
Figure (10) shows that the mean change almost fully characterizes the distribution change for Panama, that does not happen in the case of Dominican Republic shown in Figure (11). This means that, in order to have more reliable estimates for each country and each period, additional complexity is needed.

Figure 10: Income distribution of Panama



Note: the poverty line represents a household per capita income of 5.5 dollars (2017 PPP).

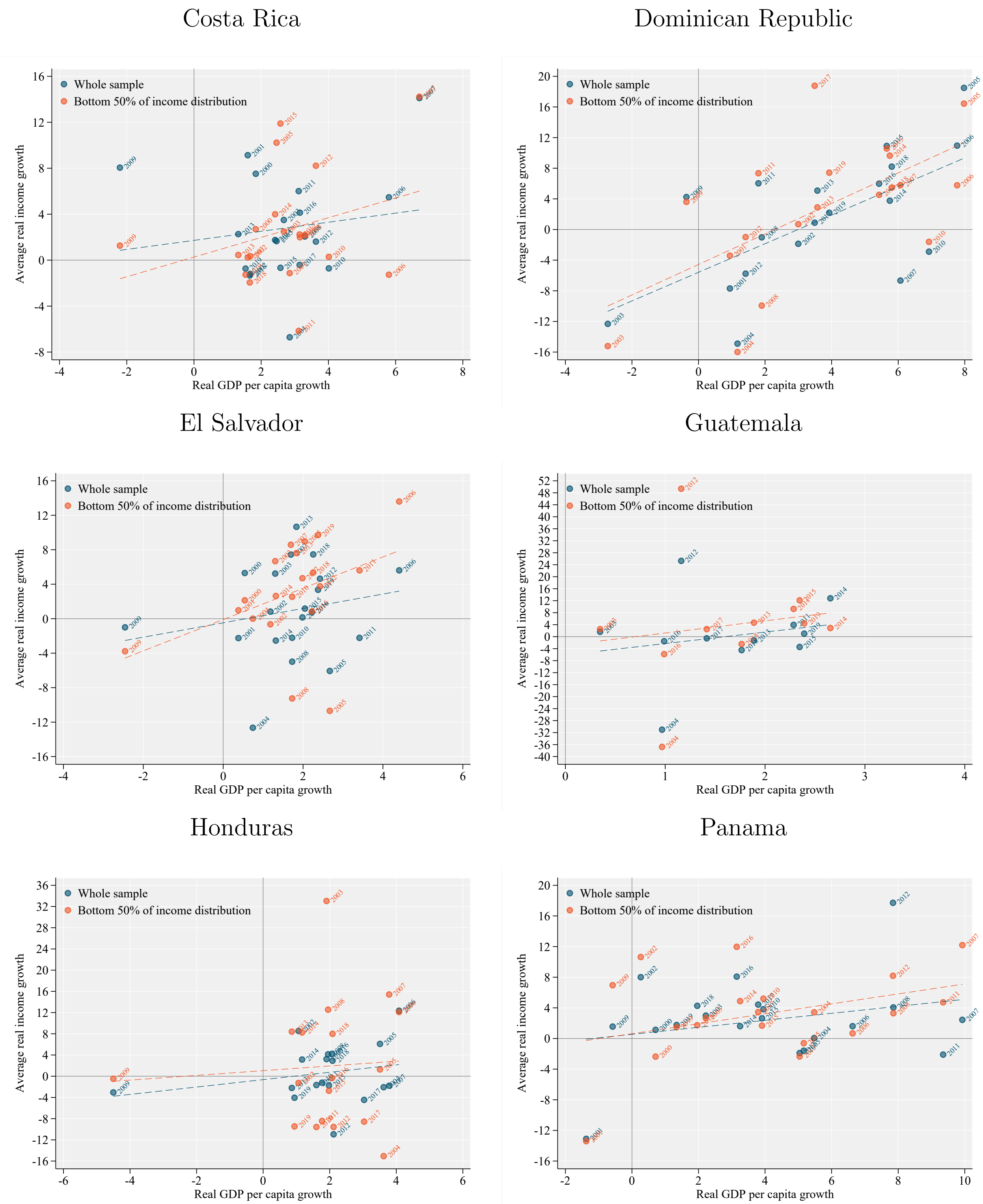
Figure 11: Income distribution of Dom. Republic



Note: the poverty line represents a household per capita income of 5.5 dollars (2017 PPP).

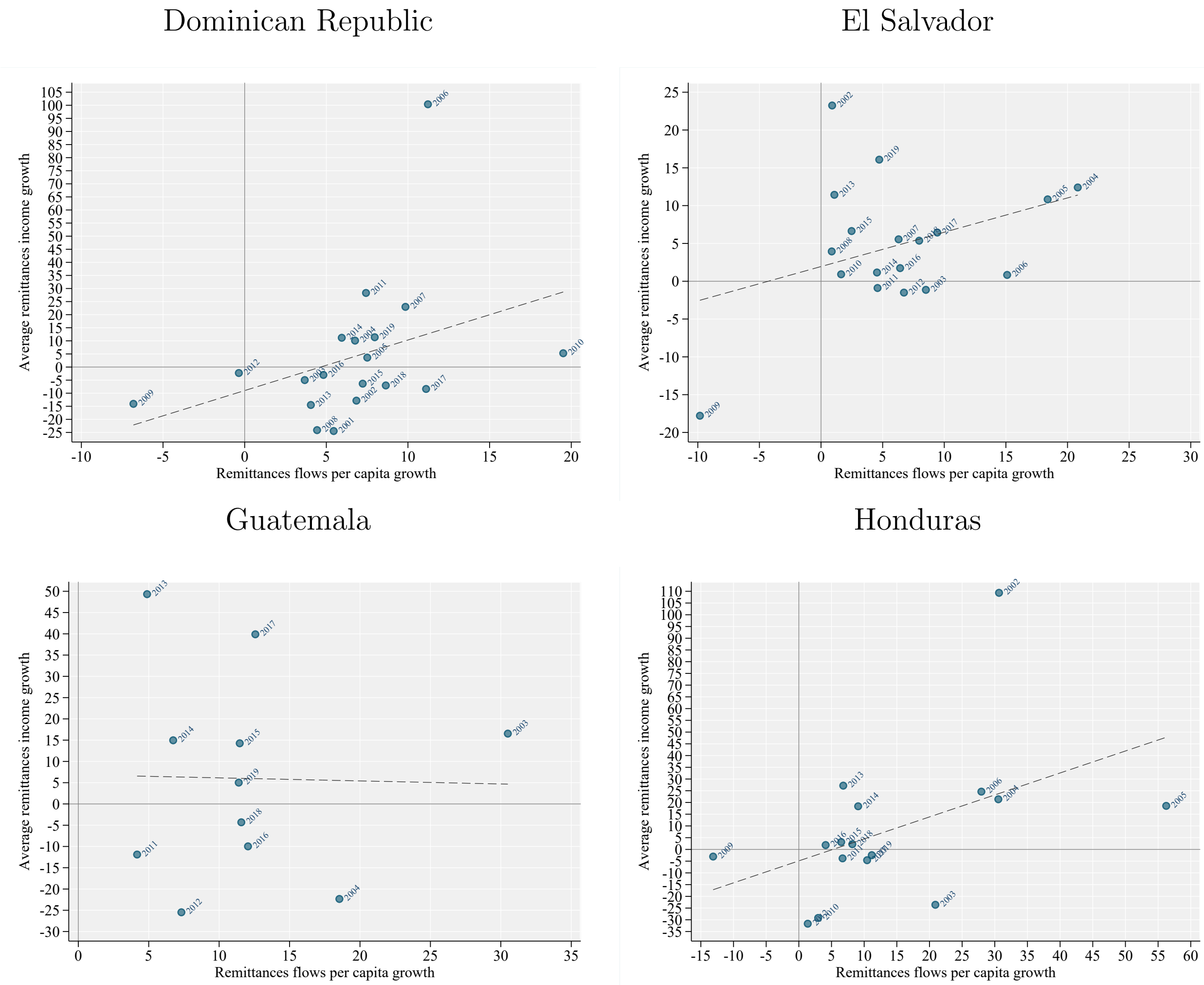
A.2

Figure 12: Changes in GDP and surveys' income



A.3

Figure 13: Changes in remittances flows and surveys' remittances income



B Additional tables

B.1

Table 3: Examples of micro-simulations

Article	Application
European Comission (2024)	Simulate tax and benefit components of household disposable income.
Bargain and Callan (2010)	Tax policy in the UK.
Molina et al. (2020)	Highlight the concept of income vulnerability as a result of the COVID-19 pandemic and poses the need (and potential) to use enriched micro-data to assess policy scenarios and distributional effects.
Rueda (2021)	Microsimulation exercise to address the effects of COVID-19 on poverty levels via income and employment shocks in Colombia. The authors also perform a policy scenario analysis on monetary transfers.
Tejerina and Muñoz (2015)	The evaluate the change over time of poverty and inequality indicators in El Salvador and the effects of a CCT program in El Salvador using a microsimulation approach.
Zegarra and Tuesta (2009)	Food safety-focused microsimulation in Peru. The dependent variable was food consumption (measure by caloric intake), and the independent variables are food prices, which, at the time, where steadily increasing
Molina et al. (2022)	Cross-country analysis of the impact of the increment of food and energy prices in poverty because of the Russia-Ukraine conflict. The perform scenario analysis and evaluate policy alternatives.
Arancibia Romero et al. (2019)	Using the structure of EUROMOD, the authors assess the effect of taxes and benefits on income distribution in Argentina, Bolivia, Colombia, Ecuador, Uruguay, and Venezuela
Nogueira et al. (2011)	Microsimulation model for Brazil which evaluates the impact of social benefits, social security contributions, and income taxes on consumption.
Larrañaga et al. (2012)	They use micro-simulation models in Chile to assess the impact on income distribution and poverty of health, pension, and taxes policy.
Cabezas and Acero (2011)	The authors simulate the effects of the program *ethical family income* in Chile in labor supply, income, inequality, and Poverty
Castañón-Herrera and Romero (2012)	Tax benefit analysis on the distributional effects of tax changes in Guatemala and, also, effects at the government revenue-level
Absalón and Urzúa (2012)	Analysis in Mexico for the effects of tax-benefits in personal income and analysis of the distributional impact of the 2010 tax reform.
Amarante et al. (2011)	Microsimulation in Uruguay aimed to determine the distributive impact of tax changes.

B.2

Table 4: Observed and estimated poverty changes

Year	CRI		DOM		GTM		HND		PAN		SLV	
	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.
2000	-2.48	0.40	.	.	.	.	.	.	0.00	-0.09	-1.64	-0.65
2001	8.79	1.17	2.01	1.33	.	.	.	.	4.03	-0.28	1.33	0.56
2002	-0.20	-0.35	-1.11	-1.38	.	.	-0.01	0.59	-3.06	-1.14	-0.33	-0.93
2003	-4.26	-0.43	7.54	8.39	-2.25	-0.26	-2.25	-1.09	-1.55	-1.44	-1.24	-0.84
2004	-0.17	0.48	8.07	13.05	10.51	0.19	1.93	0.09	0.21	-2.33	0.45	0.41
2005	-6.01	0.12	-7.90	-8.57	.	.	-2.36	-2.19	-0.23	-0.09	3.11	-0.24
2006	0.74	-1.50	-3.00	-3.94	.	.	-3.70	-1.34	-1.79	-1.53	-3.71	-0.85
2007	-5.85	-2.42	-1.61	-4.16	.	.	-4.23	-3.39	-3.01	-1.62	-3.26	-1.41
2008	2.82	-0.07	0.59	-1.56	.	.	-2.40	-1.17	-1.66	-0.49	3.34	1.24
2009	-4.33	-0.81	-1.82	-2.24	.	.	-0.72	-1.57	-0.05	-0.55	-0.03	-0.12
2010	-4.99	-0.82	-0.01	-2.09	.	.	3.27	-1.47	-2.02	-1.93	-0.23	-0.81
2011	2.70	-0.84	-1.84	-1.57	0.05	1.60	2.73	2.27	-0.92	-1.05	-0.52	-0.56
2012	-3.27	-2.30	-0.03	0.04	-13.03	-0.66	4.83	-0.17	-3.97	-1.30	-2.87	-1.13
2013	-0.22	-0.39	-0.76	-1.93	-1.60	-8.00	-1.99	-0.94	1.67	-0.05	-3.67	-1.32
2014	-0.68	-1.11	-4.05	-3.70	-2.08	1.51	-4.22	0.02	-2.24	-1.44	-0.60	0.07
2015	-0.74	-0.76	-3.93	-3.62	-0.60	-1.92	-0.61	-0.29	0.49	-1.53	-2.19	-3.61
2016	-0.71	0.43	-2.61	-2.87	0.61	0.08	-1.79	-1.80	-2.34	-1.84	-1.70	-1.57
2017	-0.48	-0.17	-4.81	-3.06	0.30	-1.26	3.02	-1.26	-0.49	-1.10	-1.48	-1.45
2018	0.79	-0.72	-2.68	-2.41	2.82	-0.03	-0.72	-0.36	-0.63	-0.87	-3.38	-2.28
2019	-0.25	-0.79	-2.53	-1.79	-1.46	-0.04	1.31	-0.20	0.31	-1.08	-2.23	-2.17
2020	8.89	4.66	8.65	4.38	.	.	.	.	.	.	3.63	4.14

Table 5: Availability of surveys

Year	CRI	DOM	GTM	HND	NIC	PAN	SLV
1990	✓	×	×	✓	×	×	×
1991	✓	×	×	✓	×	✓	×
1992	✓	×	×	✓	×	×	×
1993	✓	×	×	✓	✓	×	×
1994	✓	×	×	✓	×	×	×
1995	✓	✓	×	✓	×	✓	✓
1996	✓	✓	×	✓	×	×	✓
1997	✓	✓	×	✓	×	✓	✓
1998	✓	×	×	✓	✓	✓	✓
1999	✓	×	×	✓	✓	✓	✓
2000	✓	✓	✓	×	×	✓	✓
2001	✓	✓	×	✓	✓	✓	✓
2002	✓	✓	✓	✓	×	✓	✓
2003	✓	✓	✓	✓	×	✓	✓
2004	✓	✓	✓	✓	×	✓	✓
2005	✓	✓	×	✓	✓	✓	✓
2006	✓	✓	✓	✓	×	✓	✓
2007	✓	✓	×	✓	×	✓	✓
2008	✓	✓	×	✓	×	✓	✓
2009	✓	✓	×	✓	✓	✓	✓
2010	✓	✓	✓	✓	✓	✓	✓
2011	✓	✓	✓	✓	✓	✓	✓
2012	✓	✓	✓	✓	✓	✓	✓
2013	✓	✓	✓	✓	×	✓	✓
2014	✓	✓	✓	✓	✓	✓	✓
2015	✓	✓	✓	✓	×	✓	✓
2016	✓	✓	✓	✓	×	✓	✓
2017	✓	✓	✓	✓	×	✓	✓
2018	✓	✓	✓	✓	×	✓	✓
2019	✓	✓	✓	✓	×	✓	✓
2020	✓	✓	×	×	×	×	✓