

Inequality in Latin America Revisited: Insights from Distributional National Accounts

Mauricio De Rosa
Ignacio Flores
Marc Morgan

November 2020



WID.WORLD
THE SOURCE FOR
GLOBAL INEQUALITY DATA

Income Inequality Series for Latin America*

Technical Note

Mauricio De Rosa^{1,2}

mderosa@iecon.ccee.edu.uy

Ignacio Flores^{1,3}

ignacio.flores@insead.edu

Marc Morgan¹

marc.morgan@psemail.eu

November 2020



*We are grateful to the Statistics Division of the United Nation's Economic Commission for Latin America and the Caribbean (ECLAC), with whom we have signed an agreement for the sharing of their harmonized household surveys. These key data inputs, as well as the constant support and assistance provided by ECLAC's Statistic Division's staff, were of invaluable help to this project. We also thank the statistics division of the Peruvian tax agency (SUNAT) for preparing income tax tabulations for this study; and Catalina Galdamez Vanegas for kindly sharing Salvadorian income tax tabulations with us. We would especially like to thank Facundo Alvarado for his outstanding advice and insights throughout the whole project. We gratefully acknowledge financial support from the European Research Council (ERC Grant 856455) and the French Research Agency (EUR Grant ANR-17-EURE-0001).

^{†1}Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France.

²IECON-UDELAR, Gonzalo Ramírez 1926, Montevideo, Uruguay.

³INSEAD, Boulevard de Constance, 77300 Fontainebleau, France.

1 Introduction

This document provides an overview of the sources and methods behind the Distributional National Accounts (DINA) series presented in the World Inequality Database (WID.world, <https://wid.world>) for Latin America in the November 2020 update. Our estimates distribute total pre-tax national income to resident individuals, following the DINA guidelines (Alvaredo, Atkinson, et al., 2020). For that purpose, we combine a variety of data sources, namely harmonized household surveys, income tax records, social security registers, and national accounts. The available information allows us to study ten countries, covering around 80% of the region's population over the last two decades.

This note is a technical and summary description of our estimates. The scientific assessment of our findings, as well as a detailed analysis of several important issues raised by the challenging task of distributing national income in a situation of relatively rich but particularly imperfect data, will be presented in a separate and longer paper, currently under preparation. The results of the current update should be considered as *preliminary*, thus subject to revision and improvement. We welcome comments and feedback from the research community on this on-going project.

A lot of effort has been devoted to the harmonisation of data inputs to improve comparability. Before the publication of the present series, Brazil was the only country in the region with DINA estimates in WID.world, based on Morgan (2017). Fiscal income or rescaled fiscal income series for a number of countries such as Argentina, Chile, Colombia and Uruguay were available, but not DINA estimates accounting for the totality of national income.¹ Therefore, the November 2020 update includes brand new series for eight countries, as well as a revised and updated series for Brazil. It is expected that future updates will extend the time coverage back to 1990. Due to considerable (although not surprising) statistical inconsistencies across data sources, we do not currently present series for a number of countries, even if survey and/or administrative data exist. Such inconsistencies require more time to reconcile.

Given the complexity of the task, and the number of assumptions required for the construction of these series, this document also describes the impact that each stage of our adjustment procedure has on the income distribution, starting from the survey-based figures. We hope that this gives readers greater insight on the final series presented in the database. We distinguish three steps: first, we adjust for the low representative-

¹All working papers and technical notes are available in WID.world's library, <https://wid.world/methodology/#library-general>.

ness of top incomes in surveys using administrative records; second, we correct for the under/over coverage of income components by scaling them to national accounts aggregates (i.e., wages, capital incomes, mixed income, pensions and imputed rents); and third, we impute income from the corporate sector and the general government to the household income distribution. This allows us to fully account for the national income recorded in each country's system of national accounts.

Not surprisingly, each of the three steps of our procedure increases inequality levels, with different magnitudes depending on the country. Their impact on trends is not homogeneous either. In some countries we continue to observe the decrease in inequality during the period – such as in Colombia, Ecuador, El Salvador, Argentina and Uruguay – while in others – like Brazil, Mexico, Peru and Chile – we observe trends that gradually flatten or even revert with each step. We will address how these changes challenge the prevailing narrative from the literature in our forthcoming paper.

The note is structured as follows. Section 2 describes the sources of data; section 3 explains the methods employed to move from survey incomes to the definition of national income, as well as the unit and income concepts employed. Finally, section 4 presents the main results.

2 Data sources

We rely on four main data sources: households surveys, income tax records, social security records, and the national accounts. Table 1 schematically presents the data sources for countries included in this update, together with the years covered by each source, whereas table 2 displays data availability for countries that remain excluded for the moment. The following subsections elaborate on the databases presented in both tables.

2.1 Households surveys

We use the survey micro-data harmonized by the Statistics Division of the UN's Economic Commission for Latin America and the Caribbean (ECLAC), including ten countries for the years from 2000 to 2018: Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Mexico, Peru, and Uruguay.

ECLAC's harmonisation process builds on the original surveys produced on a yearly basis by the official statistics institutes of the countries listed in Table 1. It seeks to create comparable income variables across countries, including the decomposition in

terms of labour, capital and mixed incomes, pensions, owner occupier rental income, transfers and other incomes.² In all cases but one, post-tax incomes are recorded on an individual basis, the exception being Brazil, where pre-tax incomes are recorded. Owner occupier rental income and some capital incomes are collected at the household level, and distributed among the adults (aged 20 years and over) of the household.

Household surveys provided by ECLAC thus represent one of the key data inputs for this study. More broadly, national surveys are an extremely important reference point in their own right in Latin America, since they are the only source available in almost all the countries. Official statistics on inequality, poverty, unemployment, etc., are drawn from them. Based on ECLAC data, we are able to reproduce country level inequality estimates by the World Bank (WB), as depicted in Figure 1. This points to the fact that, even if the two harmonisation processes (ECLAC-WB) are independent, they produce very similar results in terms of income distribution³.

Table 2 makes it explicit that many of the countries that remain excluded from this update, mostly from Central America and the Caribbean, either do not report distributive data at all (Belize, Cuba, Haiti, Jamaica, Suriname and Trinidad and Tobago), do not run household surveys on a regular basis (Bahamas, Nicaragua, Venezuela), or only run surveys but do not have any kind of publicly accessible administrative data (Bolivia, Dominican Republic, Honduras, Panama and Paraguay).⁴

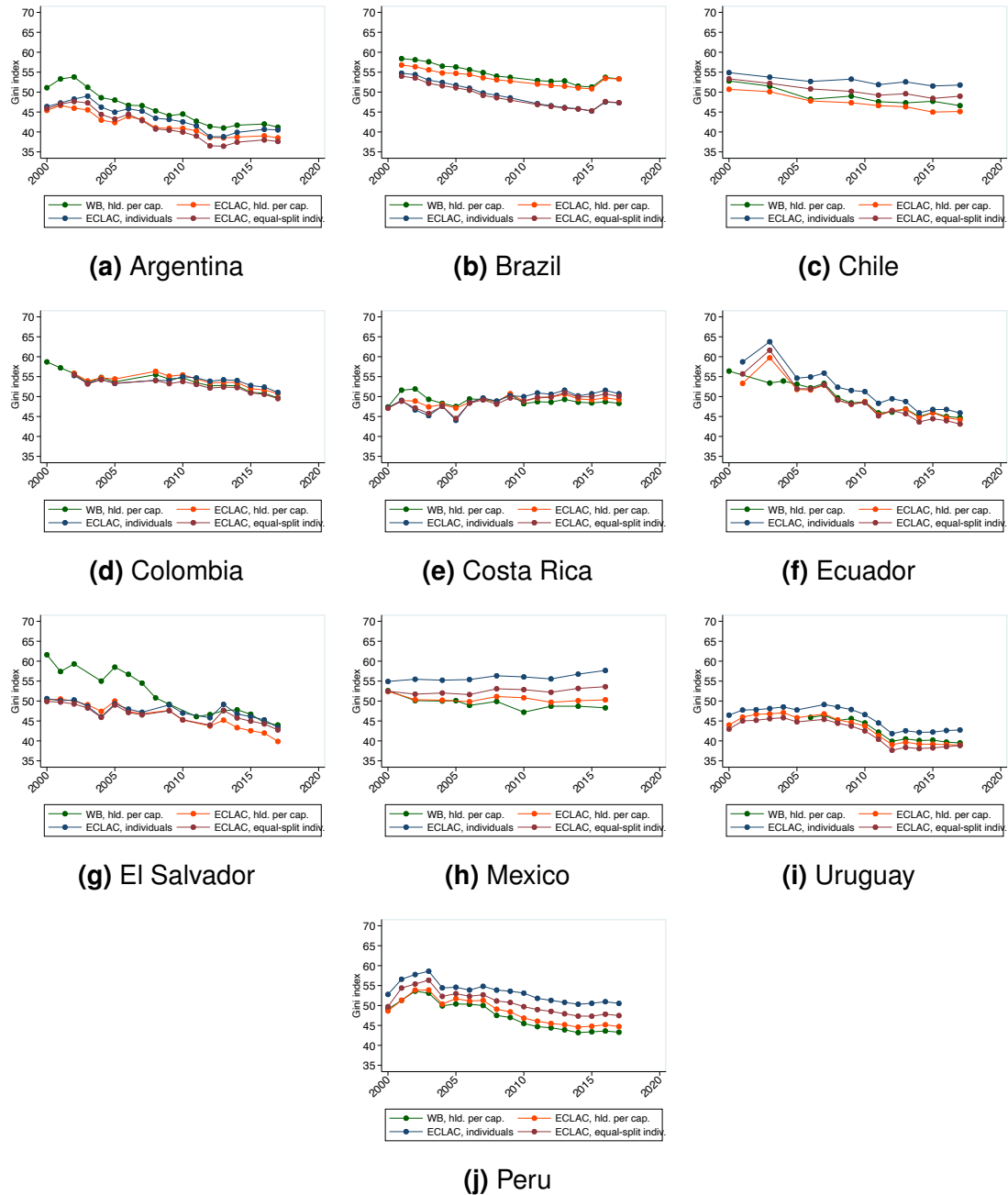
Figure 2 shows the decomposition of income in surveys, before any adjustment or correction, in terms of wages, pensions, capital income, self-employment income, and imputed rents. Wages and self-employment income represent 60-90% of total household incomes, while capital incomes are much lower.

²The only exceptions concerning the frequency of the surveys are Chile and Mexico, which collect data every two to three years.

³El Salvador up to 2010 is the clearest exception, since World Bank estimates are considerably higher and falling very rapidly. The surprisingly large inequality decrease of over 10 points in the Gini index, casts doubts on this trend, while the one resulting from ECLAC's harmonized surveys seems more reasonable.

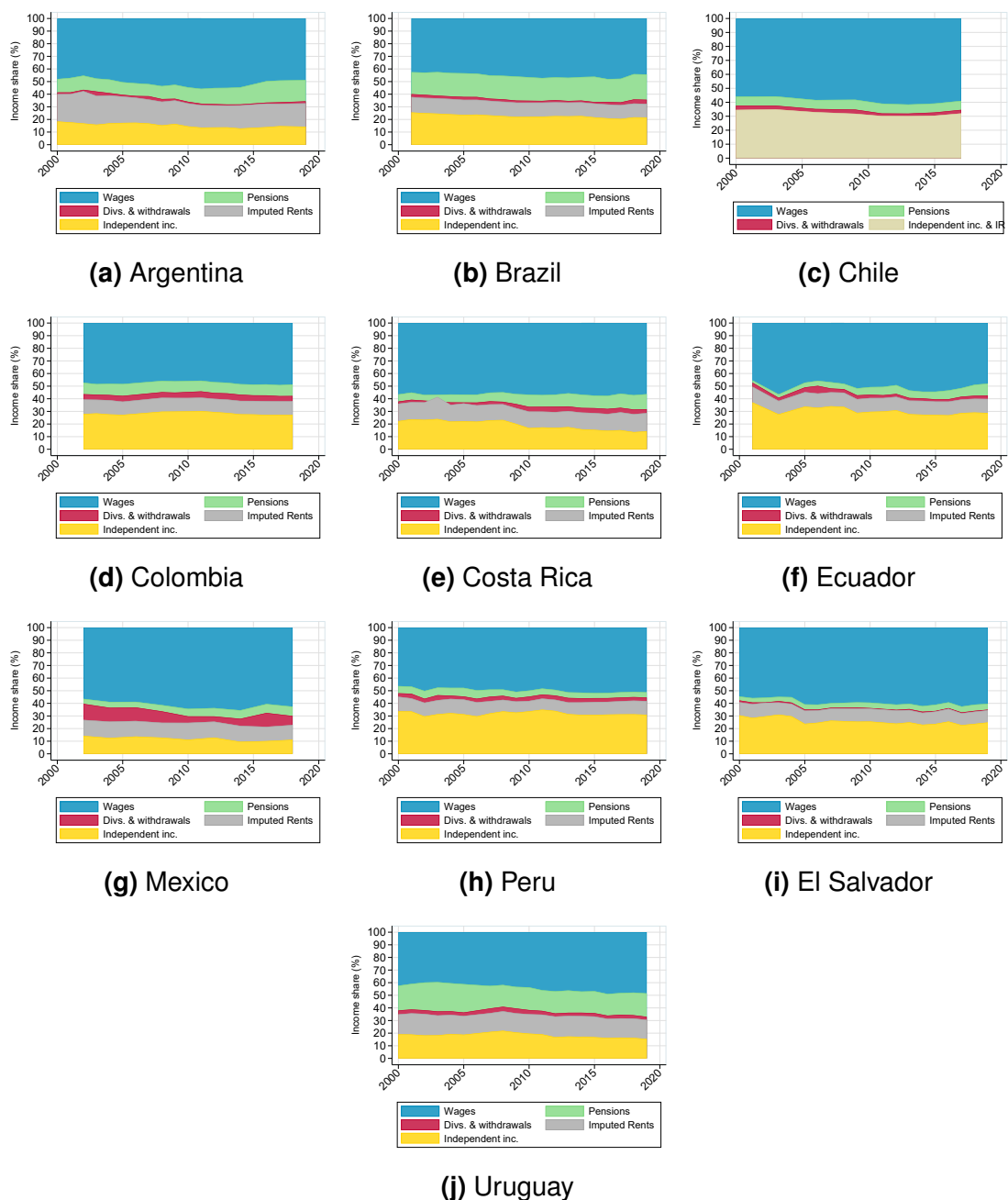
⁴In the World Inequality Database, income inequality estimates from the excluded countries of this update, are imputed based on regional averages. See Chancel and Piketty (2020).

Figure 1: Survey-based Gini indexes by source and income definition



Note. Own elaboration based on World Bank data and ECLAC's harmonized surveys. World Bank (WB) and ECLAC's household per capita income series ("hld. per cap.") show identical trends and very similar levels. The only case which presents a clear difference is El Salvador, for which World Bank's series depicts an extraordinary Gini index fall close to 20 points. Personal income Gini indices for adult population (20 and more years) based on ECLAC's harmonized surveys are also depicted along two dimensions – individual earners and equal-split individuals (where the total income of couples is divided by two). As expected, personal income series among adults show higher inequality than household per capita income series, and in some cases, such as Chile and Mexico, significantly alters inequality trends.

Figure 2: Income composition - raw survey



Note. Own elaboration based on ECLAC's harmonized surveys. Income is pretax, net of pension contributions.

Table 1: Updated countries

Country	Survey microdata		Administrative data			
	Source	Availability	Source	Availability	Population (% of total)	Exceptions - Comments
Argentina	Encuesta Permanente de Hogares (EPH) and EPH-Continua from 2003, Insituto Nacional de Estadística y Censos (INDEC)	2000-2014, 2016-2018	Income tax tabulations, Administración Federal de Ingresos Públicos (AFIP)	2000, 2018	40%	Tax data includes only wages from private sector - Survey is representative of urban areas
Brazil	Pesquisa Nacional por Amostra de Domicílios (PNAD), Instituto Brasileiro de Geografia e Estatística (IBGE)	2001-2009, 2011-2019	Income tax tabulations, Receita Federal	2000, 2002, 2006, 2007-2018	15-25%	-
Chile	Encuesta de Caracterización Socioeconómica Nacional (CASEN), Ministerio de Desarrollo Social	2000-2009 (triannual), 2011-2017 (biannual)	Income tax tabulations, Servicio de Impuestos Internos (SII)	2000-2018	~70%	Wages reported separately from other incomes in 2000-2004,
Colombia	Encuesta continua de hogares (Gran Encuesta Integrada de Hogares from 2008), Departamento Administrativo Nacional de Estadística (DANE)	2002-2005, 2008-2018	Alvaredo and Londoño Vélez (2013)	2002-2003, 2006-2010	1%	-
Costa Rica	Encuesta Nacional de Hogares, Instituto Nacional de Estadística y Censos (INEC)	2000-2019	Zuñiga-Cordero (2018)	2000-2016	~30%	Wages declared separately from other incomes in 2010-2016, only wages in 2000-2009
Ecuador	Encuesta Periódica de Empleo y Desempleo (EPED) and Encuesta de Empleo, Desempleo y Subempleo (ENEMDU) from 2003, Insituto Nacional de Estadística y Censo (INEC)	2001, 2005-2019	Cano (2015) and Rossignolo, Oliva, and Villacreses (2016)	2008-2011	1%	-
El Salvador	Encuesta de Hogares de Propósitos Múltiples, Dirección General de Estadística y Censos (DIGESTYC)	2000-2007, 2009, 2010, 2012-2019	Income tax tabulations, Dirección General de Impuestos Internos (DGII)	2000-2017	1-4%	Wages reported separately from other incomes. We only use the latter for correction.
Mexico	Encuesta Nacional de Ingresos y Gastos de los Hogares, Instituto Nacional de Estadística, Geografía e Informática (INEGI)	2002-2018 (biannual)	Income tax microdata, Servicio de Administración Tributaria (SAT)	2009-2014	~20%	Wages reported separately from other incomes
Peru	Encuesta Nacional de Hogares - Condiciones de Vida y Pobreza, Instituto Nacional de Estadística e Informática (INEI)	2000-2019	Income tax tabulations, Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT)	2016-2018	~25%	Excludes business incomes
Uruguay	Encuesta Continua de hogares (ECH), Instituto Nacional de Estadística (INE)	2000-2005, 2007-2019	Income tax microdata, Dirección General Impositiva	2009-2016	~75%	-

Note. Own elaboration.

Table 2: Excluded countries

Country	Survey microdata		
	Source	Sample size, thousands of individuals	Availability
Bahamas	Bahamas Living Conditions Survey	6	2001
Belize	-	-	-
Bolivia	Encuesta de Empleo, Desempleo y Subempleo, Instituto Nacional de Estadística y Censo (INE)	15 – 40	2000-2019
Cuba	-	-	-
Dominican Republic	Encuesta Nacional de Fuerza de Trabajo (ENFT)	15 – 30	2000-2019
Guatemala	Encuesta Nacional de Condiciones de Vida and Encuesta Nacional de Empleo e Ingresos	10 – 70	2000, 2002-2004, 2006, 2011, 2014
Guyana	-	-	-
Haiti	-	-	-
Honduras	Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM), Instituto Nacional de Estadísticas (INE)	20 – 100	2001-2018
Jamaica	-	-	-
Nicaragua	Encuesta Nacional de Hogares sobre Medición de Nivel de Vida, Instituto Nacional de Estadística y Censos de Nicaragua	20 – 35	2001, 2005, 2009, 2014
Panama	Encuesta de Hogares, Instituto Nacional de Estadística y Censo (INEC)	40 – 55	2000-2019
Paraguay	Encuesta Integrada de Hogares (EIH) and Encuesta Permanente de Hogares (EPH) from 2002, Dirección General de Estadística, Encuestas y Censos (DGEEC)	15 – 40	2001-2019
Suriname	-	-	-
Trinidad and Tobago	-	-	-
Venezuela	Encuesta de Hogares Por Muestreo (EHM), Oficina Central de Estadística e Informática	80 – 240	2000-2006

Note. Own elaboration.

2.2 Data from administrative records

Available distributional data from administrative sources in Latin America can be classified in three groups:

- (i) microdata covering those required to submit a tax file (e.g. Mexico);
- (ii) grouped data (tabulations) based on the universe of tax payers, or those required to declare their incomes, organised by ranges of income (e.g. Argentina, Brazil, Chile and Uruguay);
- (iii) distributional data covering income tax payers with wage income only, either in microdata format (e.g. Argentina, Costa Rica), or in tabulated form (e.g. Brazil);
- (iv) in an increasing number of countries, information on the distribution of wages is

made available from the social security administration, either in micro- or grouped- data format. Naturally, this is restricted to the formal sector, and, depending on each country institutional arrangements, this may include the universe of formal workers, or only those in the main social security regime. We use social security records in the case of Costa Rica.

The current update includes two completely new administrative sources for countries where –to our knowledge– tax data was never used to study the income distribution. One case is Peru, for which tax authorities kindly prepared tabulated income statistics for this study. Their data cover three years (2016-2018). It excludes business incomes, but includes pre-tax wages, dividends, interests and other incomes. The other case is El Salvador, where we gained access to two types of income tax tabulations, covering almost the whole period (2000-2017). One of the tables includes pretax wage income, while the other only includes individuals reporting income from multiple sources. However, the Salvadorian series presented in the following figures only make use of the latter tables, due to inconsistencies in the former. In Mexico we also have two different sources, but we only use wage data so far, for the same reasons. The rest of the countries in Table 1 can be divided in two groups. On the one side, those regularly publishing and updating their administrative records (Brazil, Chile, Costa Rica, Mexico and Uruguay). On the other side, those that gave external researchers access to microdata at some point, but do not produce distributive information from tax registers on a regular basis (Colombia and Ecuador). For these cases, we use estimates prepared by the authors of previous studies (Alvaredo and Londoño-Vélez, 2013; Cano, 2015; Rossignolo, Oliva, and Villacreses, 2016), which are restricted to the top percentile of the distribution only.

2.3 National Accounts

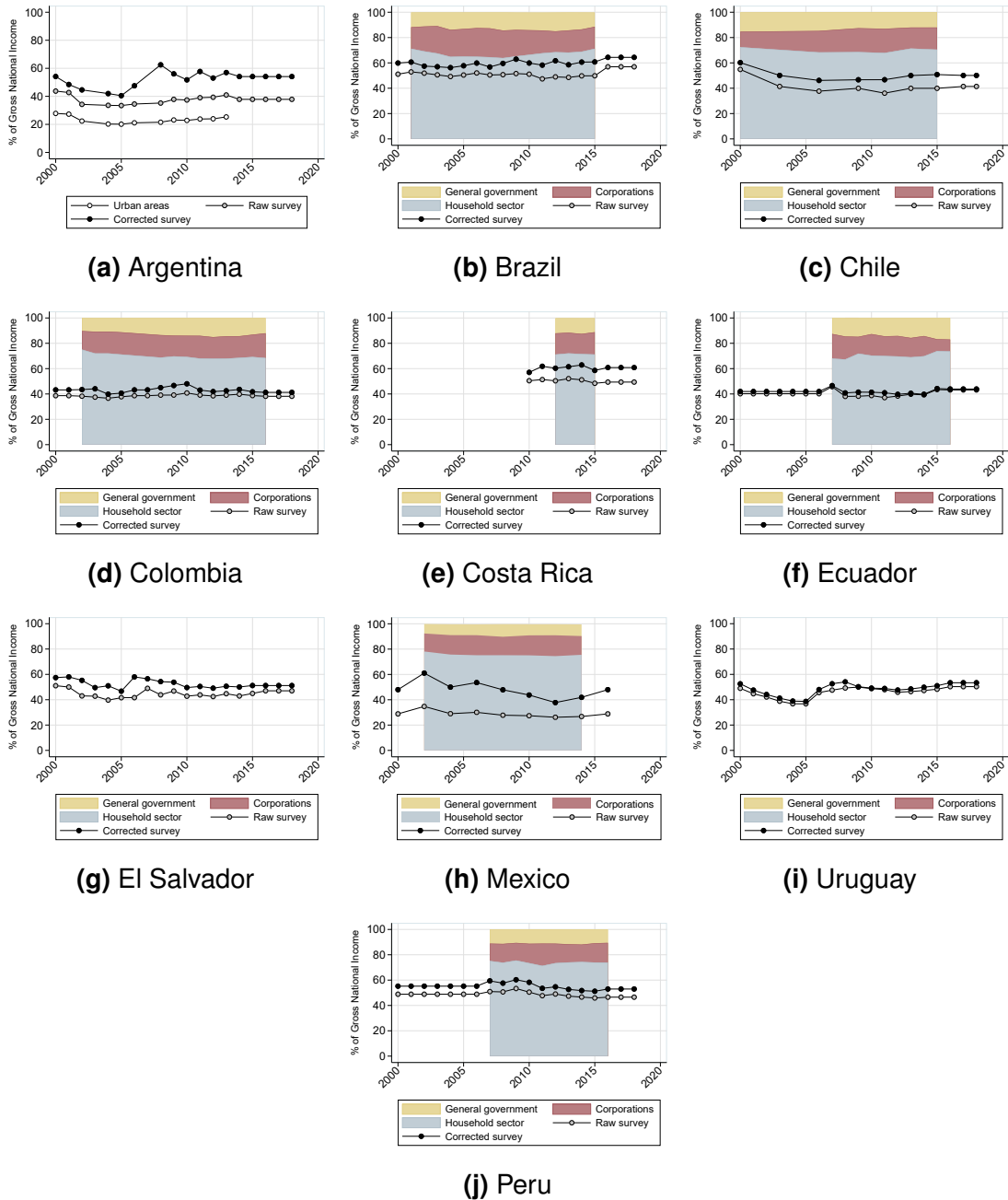
The information from National Accounts (NA) was obtained by scrapping the United Nations' Statistics Division database (<http://data.un.org>), which gathers a variety of series produced by national statistical offices. Although the macro aggregates produced by national accountants are often considered among the most reliable and internationally comparable data sources (e.g. to rank countries according to their total output, per capita GDP, etc.), detailed information on the income approach, which is the one we need to compute our series, is scarce in the region, to say the least.

Even in countries that produce this kind of data regularly, statistical agencies can update their estimates with three to five years of lag. The level of aggregation also varies across countries. For instance, despite the fact that United Nations (2009) recom-

mends distinguishing the Operating Surplus of Households (the income produced by owner-occupied housing and rented dwellings) from Mixed Income (the income of the self-employed), three countries –Chile, Ecuador and Bolivia– report both in the same aggregate. Furthermore, we observe large disparities in the level of detail provided for other relevant variables, such as the consumption of fixed capital and property incomes. This has a strong impact on our ability to account for capital depreciation, as well as to distinguish the part of the investment income of households from pension funds. These issues hinder our capacity to accurately match and compare income concepts across data sets and countries. They also force us into a trade-off between the precision of our estimates at the individual country level and their comparability at the regional level. Further comments on this matter are developed in section 3.2.

Figure 3 provides a visual comparison of aggregates across sources. It shows the decomposition of gross national income (GNI) into the household sector, the general government and the corporate sector. It also presents the aggregate income informed by surveys, before any correction, as percentage of GNI. As a preview of the results that will be discussed later, we also show the survey's total income after the adjustment with tax data. Three countries, Argentina, Uruguay and El Salvador, do not report aggregates from the income approach in the NA. For the other countries that do so, the time coverage is rather short, and usually below that of surveys. However, one result is clear: the gap between surveys, even after tax-based adjustments, and GNI is very large, usually above 40%.

Figure 3: From Household Surveys to National Income



Note. Own elaboration based on ECLAC's harmonized surveys and the UN's national accounts data. The survey series are for total pretax income. Shaded areas are the balance of primary incomes of the household sector (B.5g, S.14), corporations (B.5g, S.11 + S.12) and general government (B.5g, S.13). Point estimates of Argentina 2003 and 2007 excluded due to data inconsistencies.

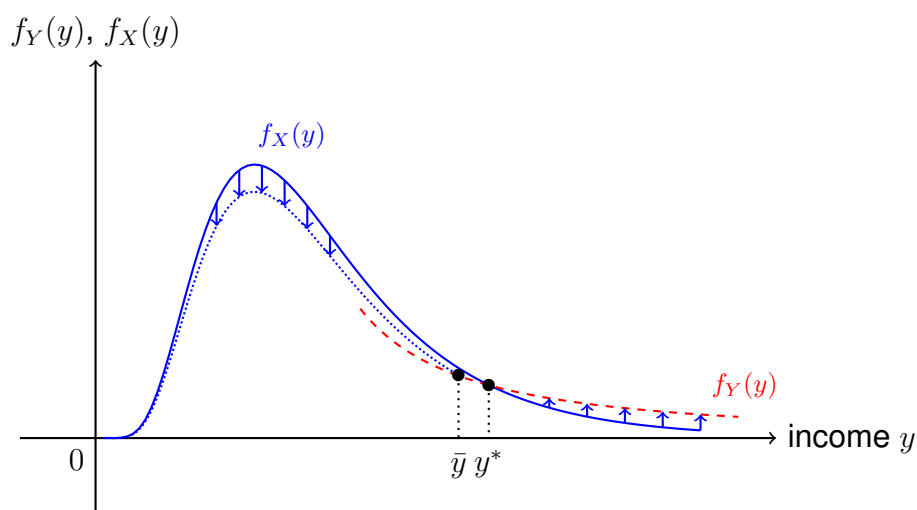
3 Estimation Methods

The transition from the survey-based distribution to the distribution of national income as measured in the National Accounts is accomplished in three steps. In the first step, we adjust household surveys to include distributive information from administrative records; in the second step, we proportionally scale the different types of income to match aggregates from national accounts; finally, in the third step, we impute corporate undistributed profits (retained earnings) and remaining missing incomes. In this section we provide a brief summary of the methods.⁵

3.1 Surveys adjusted with administrative data

The use of administrative data refers to both personal income tax declarations and social security records. These sources are mainly used to improve the coverage of top incomes groups in the survey, which are often badly captured; especially when register data is not used in the surveying process, which is the case in all countries in the region.

Figure 4: The intuition behind reweighting



Source. Blanchet, Flores, and Morgan (2018). The solid blue line represents the survey density f_X . The dashed red line represents the tax data density f_Y . Above the merging point \bar{y} , the reweighted survey data have the same distribution as the tax data (dashed red line). Below the merging point, the density has been uniformly lowered so that it still integrates to one, creating the dotted blue line.

In general, administrative records not only include individuals that are richer than the richest survey respondents, but also report bigger frequencies for moderately high

⁵For a more detailed description of Distributional National Accounts (DINA) see (Alvaredo, Atkinson, et al., 2020).

incomes. Therefore, when we compare the income distributions described in both sources, we usually find that the densities reported by administrative records tend to be higher for top incomes relative to surveys. Given that income tax declarations are made by real people, who might under-declare their income but are unlikely to over-declare, it seems natural to consider the distribution in register data as a lower bound that the survey should aim to match, at least when tax-data densities are higher.

In order to adjust the surveys we use the method described in Blanchet, Flores, and Morgan (2018), which mainly uses the ratio of survey to tax data densities to adjust survey weights. Although the method includes a “replacing” option, which allows users to impute incomes above the maximum income observed in surveys, we only use re-weighting without replacing for practical reasons (it makes the extrapolation of years without tax data clearer). The impact of not using the replacing option does not seem to affect inequality estimates in any meaningful way. Figure 4 displays the intuition behind this re-weighting process.

How do we extrapolate to years without tax data? We interpret the ratio of survey to tax densities as a rate of response, which is generally lower than one for top incomes. These are the ratios we use to adjust weights at the percentile level (with more detail in the very top). For surveys where administrative records do not exist, we assume within-country stability for these coefficients to make the adjustment.

3.2 Scaling to incomes from national accounts

Table 3: Mapping household income concepts across data sets

Income Type	National Accounts SNA08 [1] = [2] + [3]	Matching definitions SNA/Survey/Register [2]	Problematic or Missing in Survey/Register [3]
Labour	Compensation of Employees (D1)	Net wages, Salaries* (D11)	Social Security Contributions (D61)
Imp. Rents	Operating Surplus (D2)	Rent of owner occupiers	Actual rent of dwellings
Capital	Property Income** (D4)	Interest (D41), Distrib. profits (D42), Rent of natural resources (D45)	Reinvested income abroad (D43), Rent of ins. policyholders, imputed (D441), Returns on pension funds, imputed (D442), Returns on investment funds, imputed (D443)
Mixed	Mixed Income (B3)	Self employed, Independents	Rent of non-dwelling buildings
Benefits	Cash benefits (D62)	Pension benefits, Other cash benefits	

(*) Sick-leave is part of social insurance benefits in SNA, while it is part of Salaries in surveys.

(**) SNA does not deduct costs when deriving property income.

Notes: Summary table, based on United Nations, 2009 and OECD, 2013.

Figure 5 displays the adjustment factors used to scale five types of income (wages, capital incomes, mixed incomes, imputed rents, and social benefits) to corresponding aggregates from the national accounts. This is done proportionally to the amounts recorded at the individual level in the previous step (that is, proportionally to survey incomes after adjustment with administrative data).⁶ Table 3 summarises our benchmark matching of income concepts, which are gross of capital depreciation due to lack of information (we expect to subtract depreciation in subsequent versions of our series). For labour incomes, we subtract social security contributions from the compensation of employees before computing scaling factors. Since most countries' national accounts report pensions along with other benefits, we scale total benefits to that aggregate, assuming the joint distribution of pensions and other benefits is accurately described by the survey. The level of detail that is necessary to split the part of capital incomes received by pension funds is not available in most countries in the region. Therefore, we scale survey's total capital incomes to the property income (D4) aggregate in national accounts, which includes both those actually received by individuals (interests, dividends, etc.), as well as those received by insurance funds (D441), pension funds (D442), and investment funds (D443), which are imputed to households.

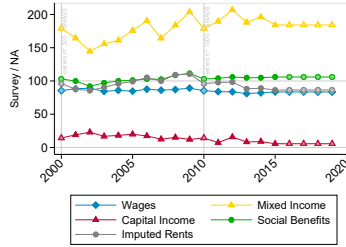
Since the income decomposition of national accounts is not available for every country and every year, we assume within-country stability of these coefficients. For countries where this decomposition is never reported (Argentina, El Salvador and Uruguay), we use the period's regional average to scale each type of income. For the final years of the period under analysis, as detailed national accounts estimates are not yet available, we extrapolated the last data point until 2019. Thus, estimates for 2016 onward might change as official data become available and are incorporated into the estimation.

⁶See Figure B.1 for a comparison of incomes from ECLAC's original surveys and corresponding aggregates from the National Accounts.

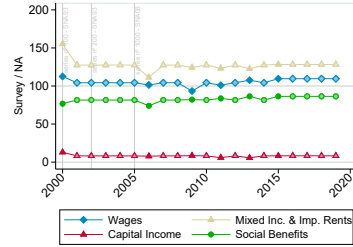
Figure 5: Scaling Factors for re-weighted Surveys



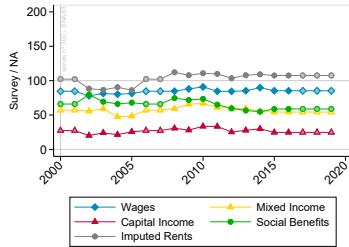
(a) Argentina



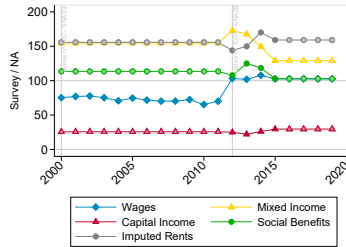
(b) Brazil



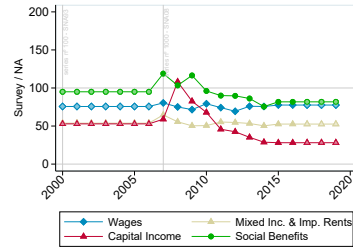
(c) Chile



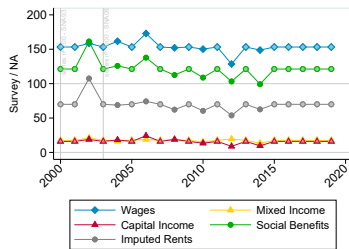
(d) Colombia



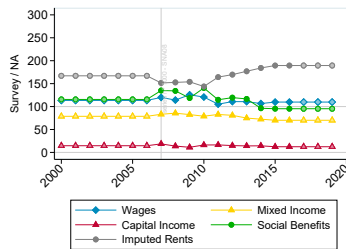
(e) Costa Rica



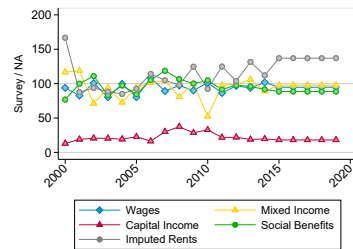
(f) Ecuador



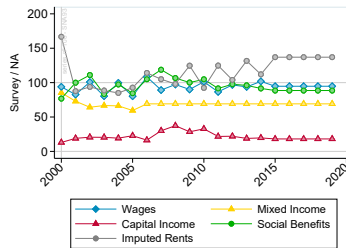
(g) Mexico



(h) Peru



(i) El Salvador



(j) Uruguay

Note. Brighter points indicate imputed scaling factors due to missing information in National Accounts. So far, we use gross estimates from national accounts because not enough countries report capital depreciation at the level that is necessary to net it out.

3.3 Imputing other missing incomes

The final step of our procedure is to impute the remaining incomes needed to reach net national income to individuals. By definition, these do not match any of the income variables that are present in the distributive data we use. Essentially, this stage boils down to the imputation of corporate undistributed profits to individuals. Since other incomes are imputed proportionally, only retained earnings have a real distributive impact. In order to estimate this aggregate, we start from the net balance of primary incomes of the corporate sector, including both the financial and non-financial sector (from WID.world). This aggregate already excludes the share of profits corresponding to portfolio investments from foreigners. In order to account for the share of undistributed profits corresponding to the government we use the share of property incomes received by the general government as a proxy (D.4 in the SNA). We thus subtract the same proportion from the net balance of primary incomes of the corporate sector. Figure 6 displays the total amount of undistributed profits, both as a share of the total income declared in the re-weighted surveys and as a share of gross national income.

In order to distribute this aggregate amount to individuals, we need a proxy for corporate ownership. Since, wealth surveys are mostly absent from the region, we use variables from income surveys as proxies. We distribute them proportionally to the sum of declared dividends, profit withdrawals and employers' income, where an employer's income refers to the total income of individuals that declare being an employer when asked about their occupation. Figure 7 depicts the incidence of this imputation across the income distribution. In general, most of it is attributed to the top quintile of the distribution, with the top 1% receiving between 30% to 60% of the total amount.

Since the amount of undistributed profits is not available for every country and every year, we proceed similarly to what was done for scaling factors, i.e. we assume within country stability of these coefficients and use regional averages for countries with no data. Moreover, data points for the final years are extrapolations of the last data point, such that results for 2016 onward should be considered as very preliminary until actual estimates for these years become available.

Figure 6: Undistributed Profits as % of Aggregate Incomes

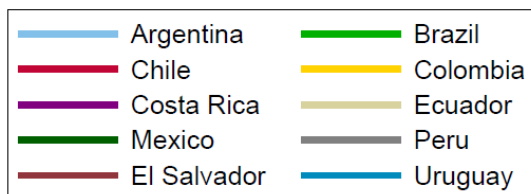
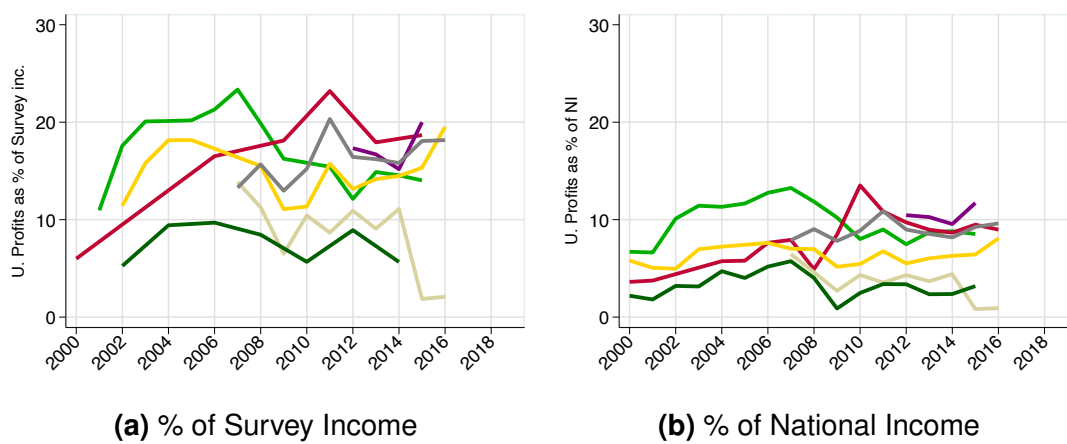
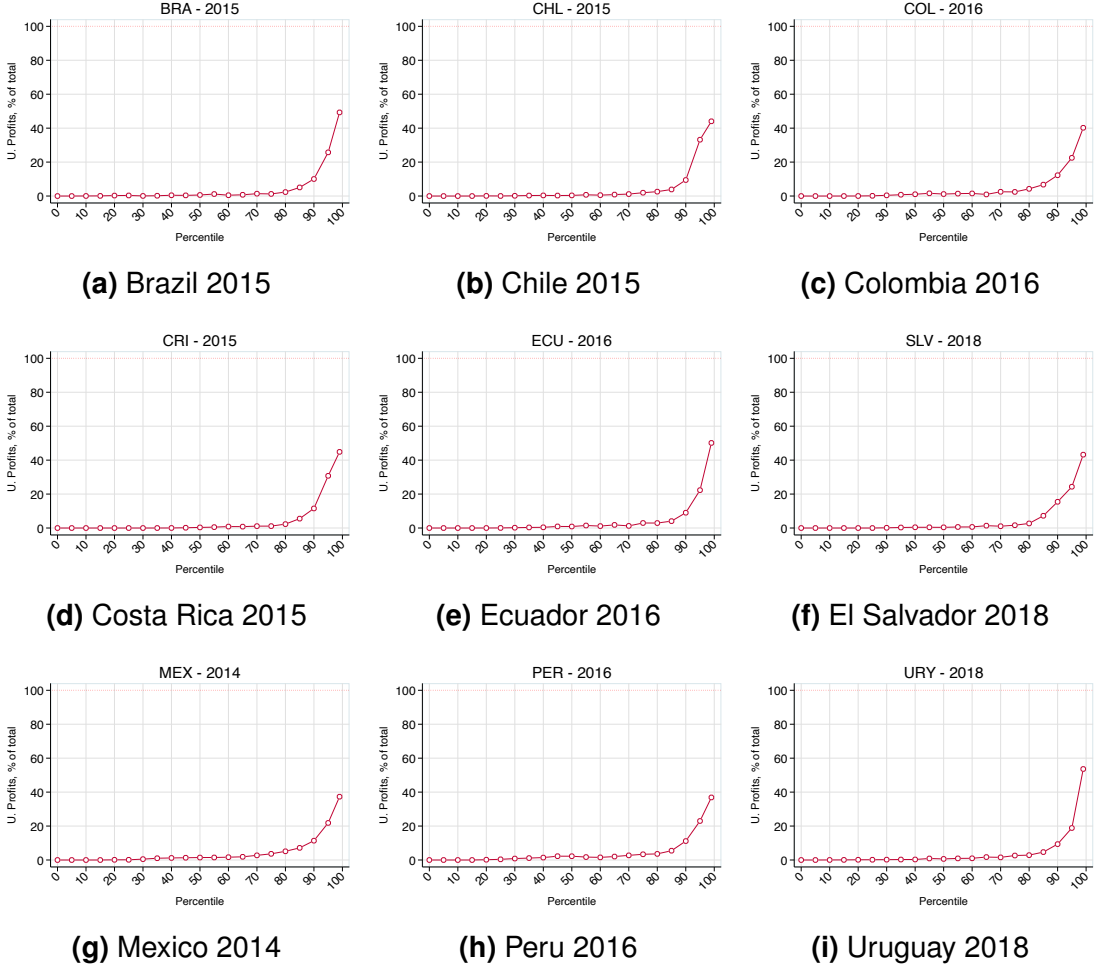


Figure 7: Share of Total Undistributed Profits Imputed to each Fractile

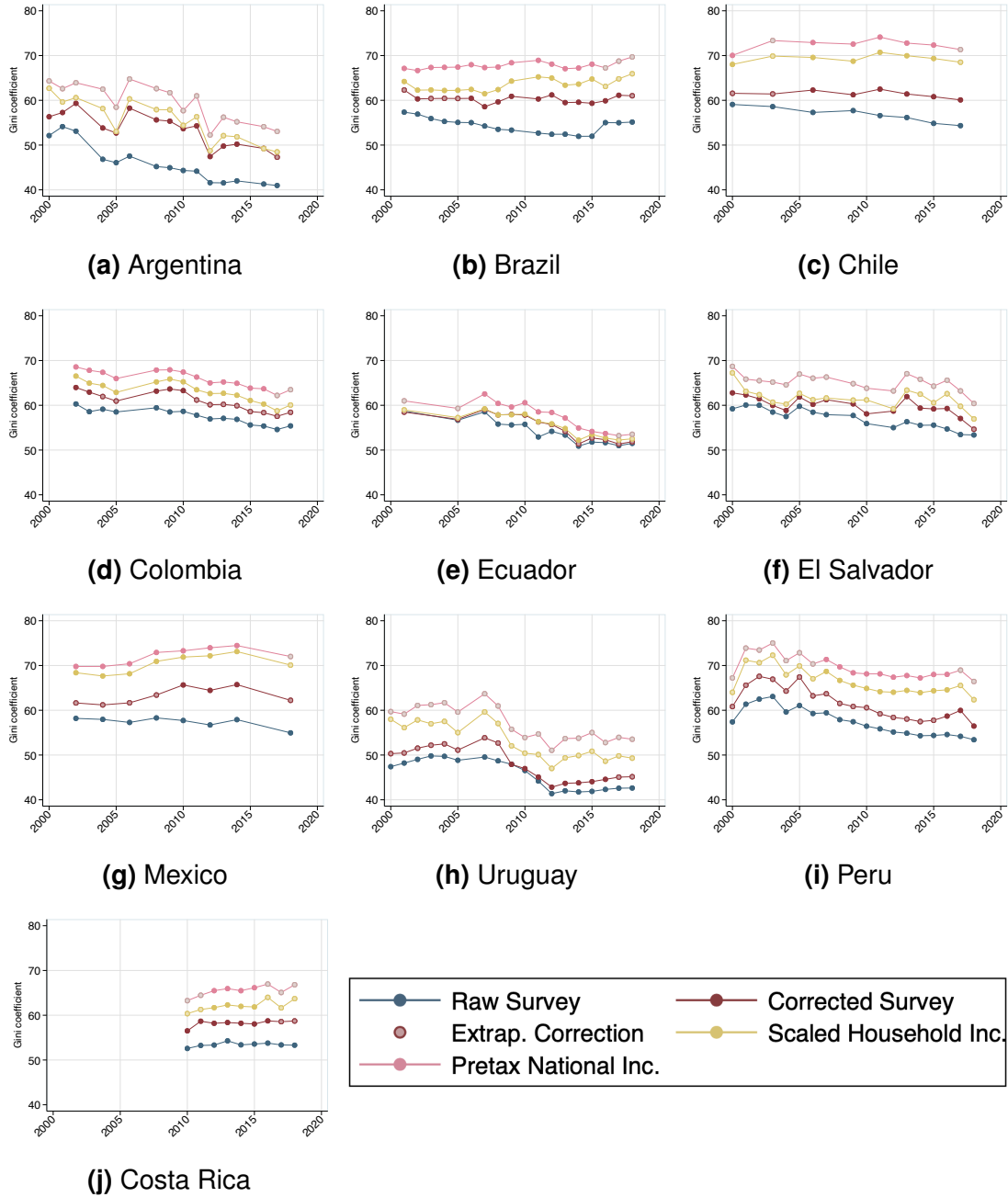


4 Results⁷

In this section we briefly present a summary of the results in graphical form. Figure 8 shows the evolution of the Gini coefficient, according to each of the adjustment steps: the original survey, the survey after adjustment with administrative data, the scaling up to NA household incomes, and the final result after the imputations that lead to pre-tax national income. Given the results on income shares shown in figures 9 to 12, it is not surprising that the Gini coefficient of pre-tax national income is higher than that of the original surveys (figure 8). Yet, there are important differences across countries, in terms of changes in levels and changes in dynamics. Our benchmark distributional estimates rank adult individuals (aged 20 and above) by increasing intervals of income, where the total income of couples is divided by two. Thus, our main series correspond to “equal-split” adult incomes (single adult incomes and equally-split couple incomes).

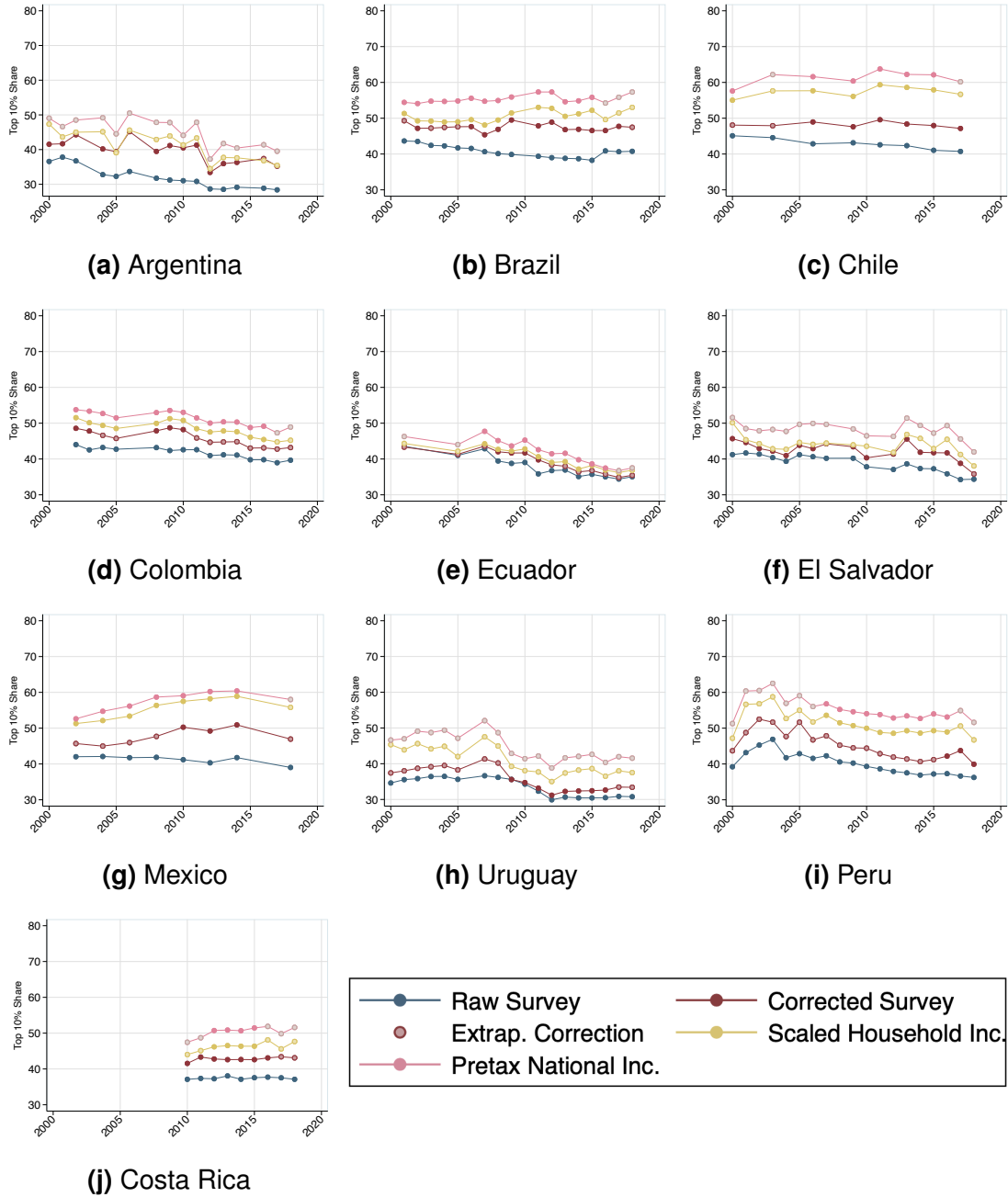
⁷The series presented on WID.world differ from the ones presented here in that they extrapolate distributional estimates to 2019 assuming a constant distribution.

Figure 8: Gini coefficients in three steps



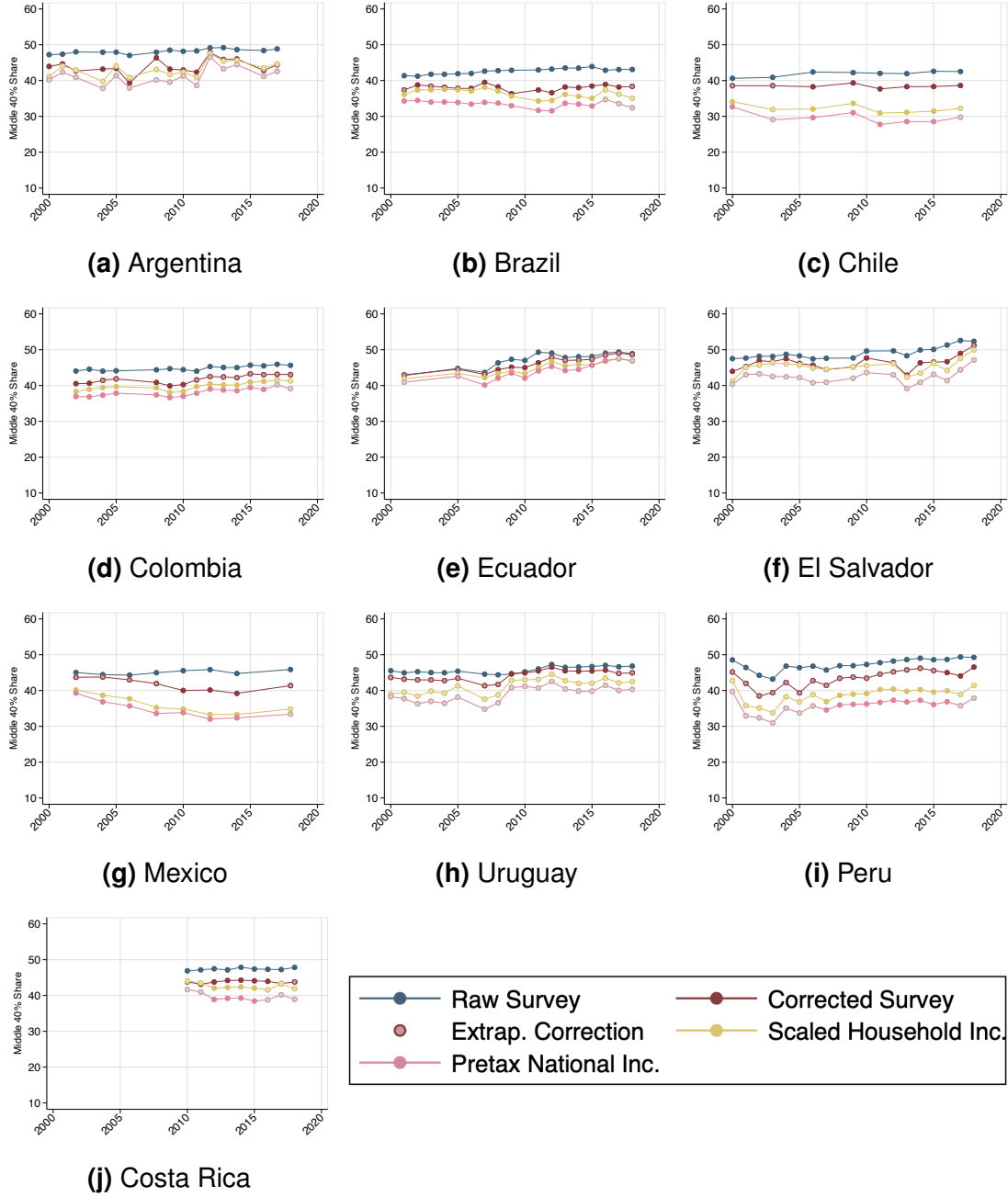
Note. Own elaboration. The figures depict household survey based estimates and the three estimation steps. The first step uses tax data to correct the raw survey; the second step scales the income totals in the corrected survey to their equivalent household-level aggregates in the national accounts; the third step imputes missing incomes needed to reach national income. Brighter points indicate that at least part of the data necessary for the stage was imputed based on remaining country/year averages. Point estimates of Argentina 2003, 2007, Ecuador 2003, 2006 and Mexico 2016 are excluded due to data inconsistencies.

Figure 9: Top 10% Share in three steps



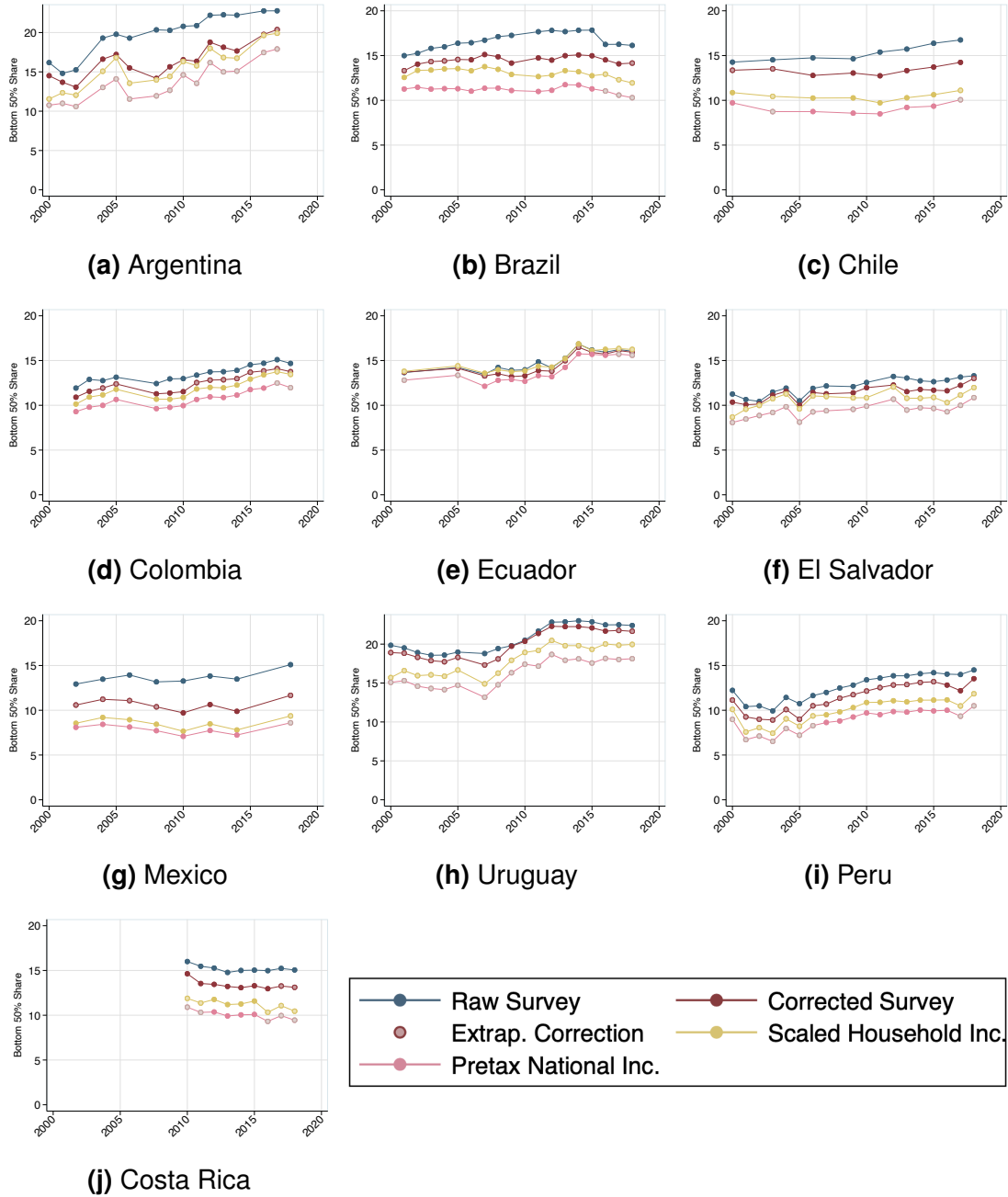
Note. Own elaboration. The figures depict household survey based estimates and the three estimation steps. The first step uses tax data to correct the raw survey; the second step scales the income totals in the corrected survey to their equivalent household-level aggregates in the national accounts; the third step imputes missing incomes needed to reach national income. Brighter points indicate that at least part of the data necessary for the stage was imputed based on remaining country/year averages. Point estimates of Argentina 2003, 2007, Ecuador 2003, 2006 and Mexico 2016 are excluded due to data inconsistencies.

Figure 10: Middle 40% Share in three steps



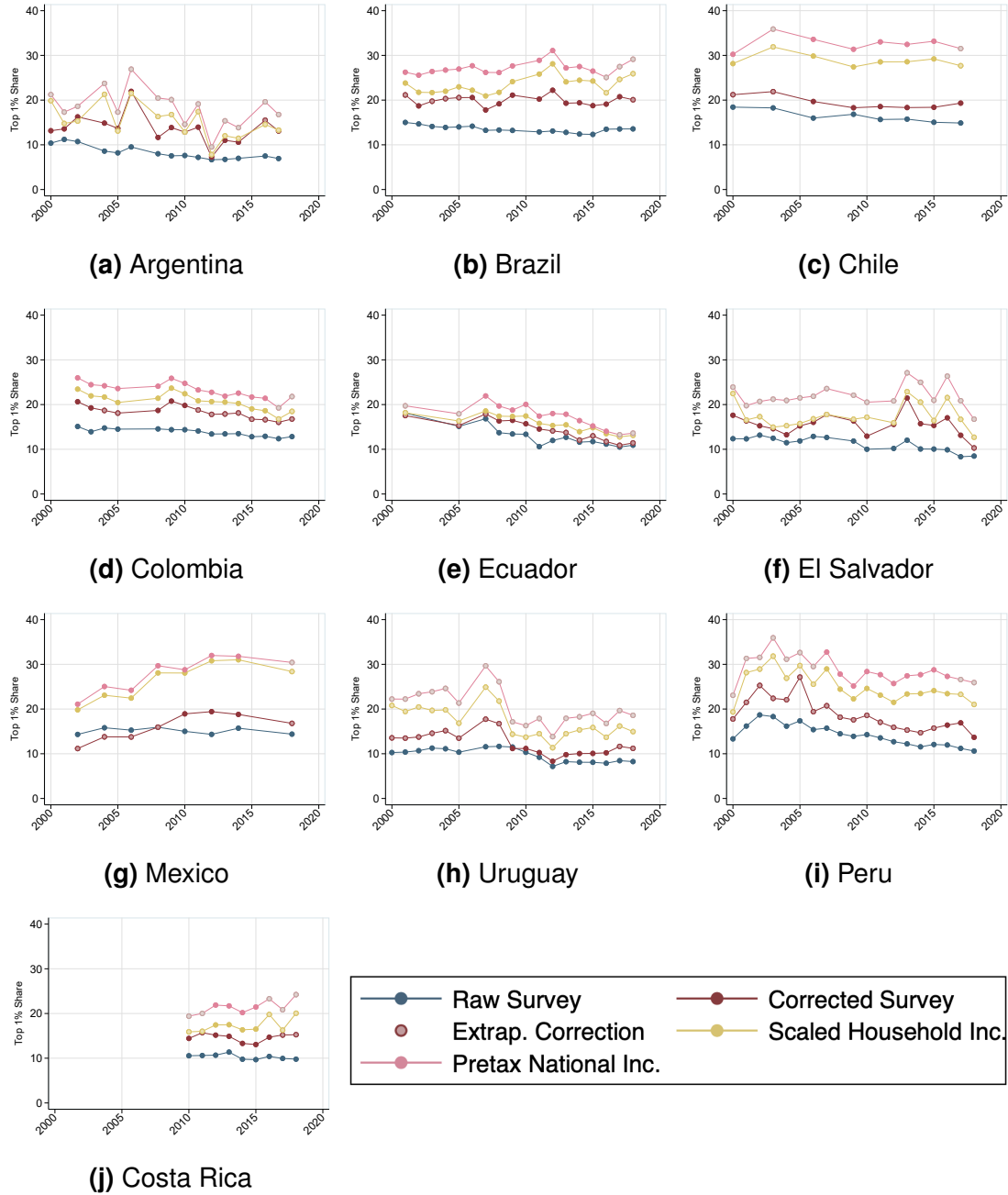
Note. Own elaboration. The figures depict household survey based estimates and the three estimation steps. The first step uses tax data to correct the raw survey; the second step scales the income totals in the corrected survey to their equivalent household-level aggregates in the national accounts; the third step imputes missing incomes needed to reach national income. Brighter points indicate that at least part of the data necessary for the stage was imputed based on remaining country/year averages. Point estimates of Argentina 2003, 2007, Ecuador 2003, 2006 and Mexico 2016 are excluded due to data inconsistencies.

Figure 11: Bottom 50% Share in three steps



Note. Own elaboration. The figures depict household survey based estimates and the three estimation steps. The first step uses tax data to correct the raw survey; the second step scales the income totals in the corrected survey to their equivalent household-level aggregates in the national accounts; the third step imputes missing incomes needed to reach national income. Brighter points indicate that at least part of the data necessary for the stage was imputed based on remaining country/year averages. Point estimates of Argentina 2003, 2007, Ecuador 2003, 2006 and Mexico 2016 are excluded due to data inconsistencies.

Figure 12: Top 1% Share in three steps



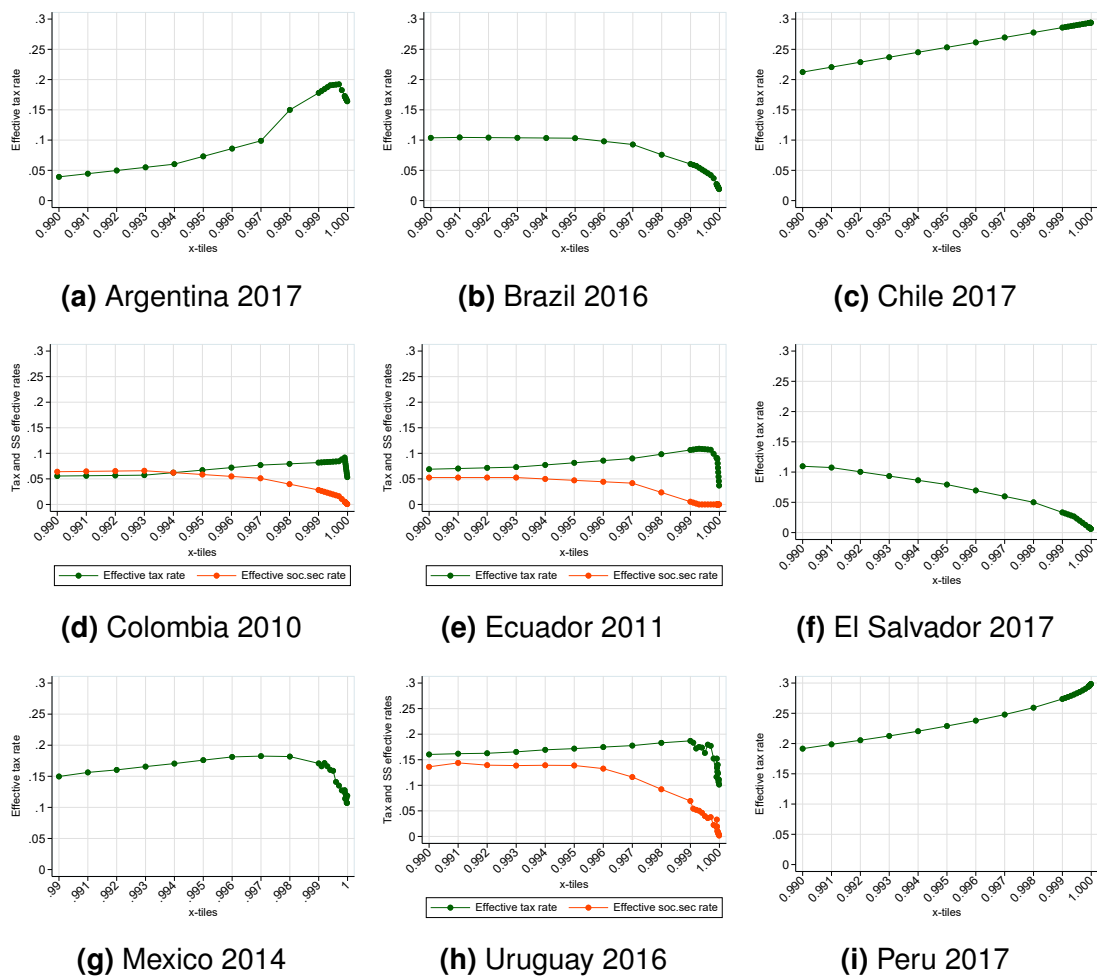
Note. Own elaboration. The figures depict household survey based estimates and the three estimation steps. The first step uses tax data to correct the raw survey; the second step scales the income totals in the corrected survey to their equivalent household-level aggregates in the national accounts; the third step imputes missing incomes needed to reach national income. Brighter points indicate that at least part of the data necessary for the stage was imputed based on remaining country/year averages. Point estimates of Argentina 2003, 2007, Ecuador 2003, 2006 and Mexico 2016 are excluded due to data inconsistencies.

Appendix

A Estimation of pre-tax income

The benchmark inequality estimates in the DINA framework is the pre-tax National Income series. However, the main data-source on which our estimates are based are harmonized household surveys, which account for household's post-tax income⁸. Thus, both for the tax-data adjustment of survey's based on Blanchet, Flores, and Morgan (2018), and for the scaling up to National Income process, it is necessary to calculate pre-tax incomes at the survey level.

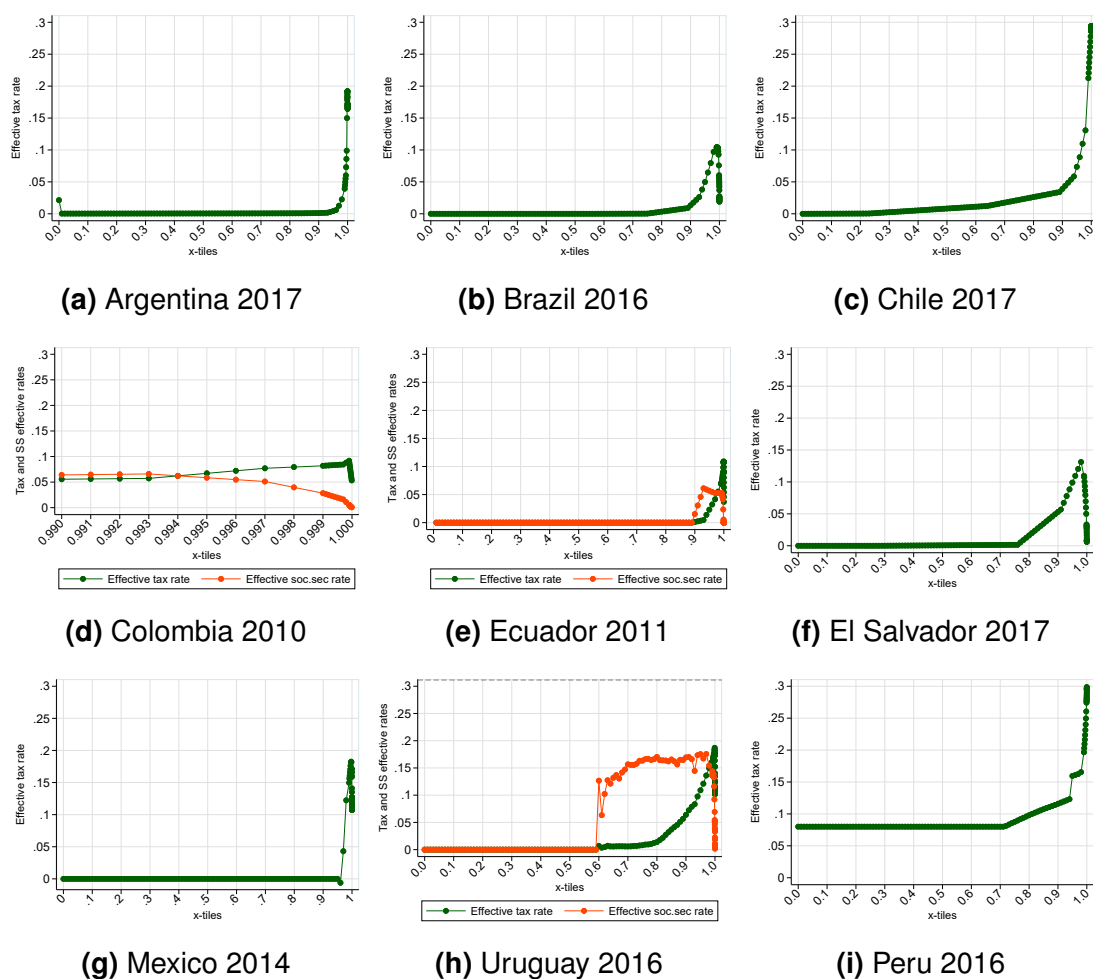
Figure A.1: Effective tax and social security rates - Top 1% - Latest year



As data on direct taxes paid by individuals or households is not collected in surveys, in order to estimate pre-tax incomes we consider external sources, mainly tax data. Broadly speaking, we compute effective tax rates by income fractile in the tax data, and use these tax rates to calculate pre-tax incomes in the survey, based on the income

⁸The only exception is Brazil, whose survey accounts for pre-tax income.

Figure A.2: Effective tax and social security rates - Latest year



fractiles to which individuals belong to.⁹ Effective tax rates by income fractile are computed for the years for which we have access to this data-source, and the average effective tax rate by fractile is used to calculate pre-tax incomes when tax data is not available.¹⁰ Tax data quality and coverage, however, varies significantly across countries and so specific procedures and assumptions have to be made for each country. In Table A.1, the main characteristics of the data and estimation procedure by country are depicted.

In the cases where available data comes from tax tabulations, effective rates are computed for observed points (e.g. the average of a given income bracket) and linearly interpolated. For Colombia and Ecuador, effective tax rates are taken directly from other studies – Londoño-Vélez, 2012 for Colombia or Cano (2015) and Rossignolo, Oliva, and

⁹We consider, whenever possible, 127 income fractiles, which account for the whole income distribution (the first 99 percentiles) and a very detailed break-down of the top 1%, where tax rates may experience significant changes.

¹⁰This assumption is potentially problematic in the cases for which the absence of tax data reflects the absence of progressive income taxation (e.g. Uruguay prior to 2009), or when the availability of data followed a large tax reform.

Table A.1: Effective tax rates estimation by country

Country	Period	Pop. Cov.	Data	Method	Ref. income	Rates
Mexico	2009-2014	Top 2%	Microdata	Directly computed	Gross income	Tax rate
Argentina	2002-2017	Universe	Tabulations	Interpolated	Gross income?	Tax rate
Brazil	2008-2016	Universe	Microdata	Directly computed	Net income	Tax rate
Colombia	2006-2010	Top 1%	Tabulations	Interpolated*	Gross income	Tax & SS rate
Chile	2005-2017	Universe	Tabulations	Interpolated	Net income?	Tax rate
El Salvador	2000-2017	Universe	Tabulations	Interpolated	Gross income	Tax rate
Uruguay	2009-2016	Universe	Tabulations	Directly computed	Gross income	Tax & SS rate
Peru	2016-2017	Universe	Tabulations	Interpolated	Net income?	Tax rate
Ecuador	2008-2011	Top 10%	Tabulations	Interpolated*	Gross income	Tax & SS rate

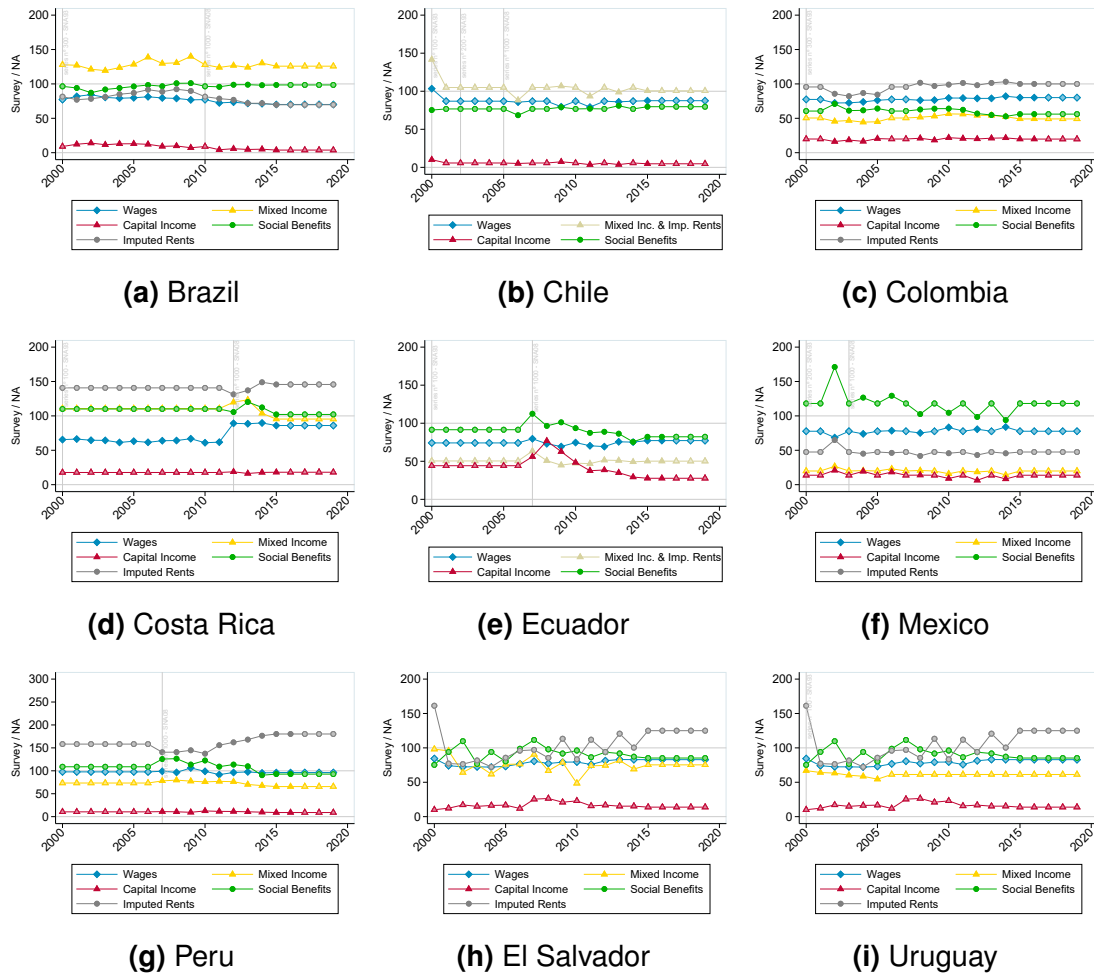
Note. Own elaboration.

Villacreses (2016) for Ecuador. Finally, for countries in which we have tax micro-data or very detailed tabulations, the effective tax rates were computed directly (e.g. Mexico and Uruguay).

Taxes are progressive, but effective rates decrease significantly in the (far) right tail of the distribution for most countries. In countries where this is not the case (Argentina) However, we cannot observe the very high income fractiles in the data without extrapolating. When social security contributions are observed (Colombia, Uruguay and Ecuador), they are a lot more regressive than the income tax, especially for top fractiles, where it converges to zero as a result of truncated schedules (i.e. schedules where a maximum income is defined for contributions).

B Comparing aggregates in National Accounts' and ECLAC's Raw Surveys

Figure B.1: Scaling factors, Raw Surveys



Note. Brighter points indicate imputations due to missing information in National Accounts.

References

- Alvaredo, Facundo, Anthony B. Atkinson, et al. "Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database". 2020.
- Alvaredo, Facundo and Juliana Londoño-Vélez. "High Incomes and Personal Taxation in a Developing Economy: Colombia (1993-2010)". 2013.
- Blanchet, Thomas, Ignacio Flores, and Marc Morgan. "The Weight of the Rich: Improving Surveys with Tax Data". 2018.
- Cano, Liliana. "Income mobility in Ecuador new evidence from individual income tax returns". Helsinki, 2015.
- Chancel, Lucas and Thomas Piketty. "Countries with Regional Imputations on WID.world: A Precautionary Note". 2020.
- Londoño-Vélez, Juliana. "Income and Wealth at the Top in Colombia: An Exploration of Tax Records 1993-2010". Master's Dissertation. Paris: Paris School of Economics, 2012.
- Morgan, Marc. "Falling Inequality beneath Extreme and Persistent Concentration: New Evidence for Brazil Combining National Accounts, Surveys and Fiscal Data, 2001-2015". 2017.
- OECD. *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*. OECD Publishing, 2013.
- Rossignolo, Darío, Nicolas Oliva, and Nestor Villacreses. "Cálculo de la concentración de los altos ingresos sobre la base de los datos impositivos: un análisis para el Ecuador". Santiago, 2016.
- United Nations. *System of National Accounts 2008*. 2009.