

More Unequal or Not as Rich?

On the Missing Half of Latin American Income*

Mauricio De Rosa[†] Ignacio Flores[‡] Marc Morgan[§]

February 28, 2021

Abstract

Since the beginning of this century, official estimates recorded a wide fall of income inequality in Latin America, which happened in parallel with strong economic growth. This gave reason to believe that a growing share of the economy was being more equally distributed. Such a trend, which is exclusively based on household survey data, was received with optimism as it occurred in a region with historically high levels of inequality. Because the opposite tendency was the norm elsewhere, it was also recognized as an exception. However, studies using administrative data combined with national accounts showed that top income shares remained relatively stable at exceptionally high levels throughout the period, contrary to the evidence from surveys. In this context, can we build a consistent and comprehensive narrative about inequality in the region? Can we use the strengths of some data sources to adjust for the weaknesses of others? This article contributes many steps in this direction. We combine surveys –which only account for roughly half of national income– with other sources for ten countries to critically examine different resulting distributions. We find that the downward inequality trend ceases to exist in 40% of our sample, even before making the estimates macro-consistent. We conclude, after presenting a range of estimates, that the region must be more unequal or not as rich as previously thought.

Keywords: Income Distribution; Macroeconomic Consistency; Survey Adjustment.

JEL Codes: C81; D31; E01.

*We are grateful to the Statistics Division of the United Nation’s Economic Commission for Latin America and the Caribbean (ECLAC) for sharing their harmonized household surveys. 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 gratefully acknowledge financial support from the European Research Council (ERC Grant 856455), the French Research Agency (EUR Grant ANR-17-EURE-0001), and the Institute for New Economic Thinking.

[†]Paris School of Economics, World Inequality Lab, Instituto de Economía - Universidad de la República (mderosa@iecon.ccee.edu.uy).

[‡]INSEAD, Paris School of Economics, World Inequality Lab (ignacio.flores@insead.edu).

[§]Paris School of Economics, World Inequality Lab (marc.morgan@psemail.eu).

Introduction

Paradoxically, during the age of ‘big data’, major obstacles impede comprehensive assessments of the income distribution both within and between countries. In particular, a prevailing problem is that different data sources used to study this issue –namely, household surveys, tax declarations, social security records and national accounts– often present remarkable discrepancies across income components. This is not only due to measurement error, but also to conceptual differences. Similarly, for a given income concept, different sources will generally describe different distributions, or at least different sections of them. Such a configuration hinders our capacity to answer surprisingly simple questions. How is economic growth distributed within countries? What would the level of inequality be, and how would it evolve, if we accounted for total macroeconomic income, instead of only the income declared in surveys? What do we know about the missing portion of income? If included, would the inequality ranking of countries or regions remain unchanged?

To answer these questions, distributional estimates must be reconciled with macroeconomic aggregates, which follow homogeneous definitions across countries. Despite considerable efforts to define a benchmark for macro-consistent inequality estimates, leading initiatives have mostly focused on developed-country contexts so far, overlooking issues that are particular to a majority of developing countries (Alvaredo et al., 2016; Zwijsenburg et al., 2017; Zwijsenburg, 2019). For instance, the two pioneering studies of Distributional National Accounts (DINA) in the United States and France heavily rely on tax micro-data to portray income distributions (Piketty et al., 2018; Garbinti et al., 2018); only using surveys to describe the very poor. In developing countries, tax coverage and compliance are much lower. In such a setting, tax data can only be used to portray top deciles, while household surveys can inform on middle and bottom incomes, which generally have higher informality rates. The challenge is thus to develop tailored methods to describe income distributions that account for different data contexts.

The first contribution of this paper is to use all available data to build a pioneering set of macro-consistent inequality estimates in Latin America –a highly unequal region whose economies have registered strong growth since the turn of the century. We present distributional results for 10 countries (Argentina, Brazil, Chile, Costa Rica, Colombia, Ecuador, Mexico, Peru, El Salvador and Uruguay), covering close to 80% of the regional population over the last two decades. We combine harmonized survey microdata from the United Nation’s Economic Commission for Latin America and the Caribbean (UNECLAC), with administrative data from tax records, social security registers and national macroeconomic accounts. In doing so we make use of recent methodological improvements to combine surveys and administrative data (Blanchet, Flores, and Morgan, 2018). We reconcile all available distributional data to build inequality estimates that not only adjust for surveys’ measurement biases, but also include imputed incomes to ensure

macroeconomic consistency. The estimates following our adjustments end up doubling the total income originally declared in most surveys. The underestimation varies considerably by income component, being more severe for non-labor incomes. This level of micro-macro discrepancy implies that our results strongly depend on methodological choices and assumptions. In order to ensure transparency, we show the impact that each step of our methodology has for the resulting distributions. We distinguish four distributional steps: first, we estimate the distribution of income in the harmonized survey data; second, we adjust for the low representativeness of top incomes in surveys using administrative records; third, we scale the main income components to their matched national accounts aggregates (i.e., wages, capital incomes, mixed income, pensions and imputed rents); and fourth, we impute remaining income items not flowing to the household sector in the national accounts (corporate retained earnings and other incomes) to reach the national income of the total economy.

Although the reconciliation of micro and macro estimates of income is a relatively new and important topic for the developed world (the next revision of the United Nation's System of National Accounts intends to incorporate it into its guidelines), this is not a new topic in Latin America. Following Altimir (1987), macro-adjustments to the inequality series in Latin American countries were made by the UN-ECLAC during more than two decades, before being discontinued for data reasons. Our work seeks to build on this 'lost tradition' in the region.

Our second contribution is to reassess the stylized facts about Latin American inequality and growth in the light of our results. Regarding the *level* of inequality, our findings suggest that the possible answers to the question posed in the title are mutually-exclusive. If we assume that the national accounts are an accurate benchmark, and proceed to distribute net national income of the total economy, our conclusion is that inequality is in fact much higher than previously thought. After adding administrative incomes to the surveys and scaling income components to the national accounts, inequality levels increase significantly –the Gini coefficients in our sample increase by about 10 points, with notable heterogeneity by country. If, on the other hand, we assume that official surveys are the benchmark for household incomes, our results are consistent with the current consensus. However, one would also need to accept that Latin American households are considerably poorer than what is reported by official macroeconomic statistics.

The analysis of inequality *trends* is not as straightforward. The adjustments we make to the survey distribution are enough to cancel out the inequality decline in countries where it was present –Brazil, Mexico, Peru and Chile– or to increase inequality where it was stable –Costa Rica. In the remaining countries (Argentina, Colombia, Ecuador, El Salvador, and Uruguay) the falling inequality trends persist even after our three sets of adjustments. However, most of the trend-reversing action operates before we scaling micro-incomes to their macro aggregates from the national accounts, i.e. after the tax-data

adjustment. In this case, there is room to believe that both statements may be true: inequality did not fall as the prevalent narrative says it did, even if countries are not as rich as what is estimated by national accountants. Thus we stress caution in proclaiming definitive statements for the region given persistent data limitations, which we flag below.

The remainder of the paper is structured as follows. In section 1 we relate our paper to the existing literature. In section 2 we describe the main datasets we use, while section 3 lays out the methodological framework. This is followed by the presentation of our distributional estimates in section 4, before concluding.

1 Literature

This section reviews two strands of literature: first, we take stock of stylized inequality facts in Latin America; second, we discuss recent projects aiming to ensure the macroeconomic consistency of distributional estimates.

1.1 Previous Work on Latin American Inequality

Latin America has a long history of inequality measurement and analysis. In this section we refer to regional studies that have contributed to our understanding of inequality dynamics in the past few decades. Numerous studies have documented and explained the apparent decline of income inequality taking place in the majority of Latin American countries during the 2000s (for examples see López-Calva and Lustig (2010); Cornia (2014); Rodríguez-Castelán et al. (2016); Messina and Silva (2017)). This decline has been viewed as exceptional in historical terms (Bértola and Williamson, 2017). For example, Cornia (2014) documents that the average regional decline in the Gini index was 5.5 points from 2002-10, after two decades of systematic increases. However, after noting that conventional data sources are not able to properly account for capital incomes or labor incomes of the “working rich”, the author argues that the evolution in 1990-2010 was driven by wage income inequality, matched by skill premium shifts benefiting the bottom of the distribution. The central role played by wage inequality in the Latin American inequality reduction is also emphasized in López-Calva and Lustig (2010) and Messina and Silva (2017). However, this decline seems to have stopped, and even reversed in some countries, in the last five years (Gasparini et al., 2018).

Tax data have seldom been integrated into the picture nor have findings from this literature been reconciled with the “conventional wisdom” in a systematic way. Where it has been attempted, the conclusion reached is that the conventional wisdom regarding inequality trends remains solid. For example, De la Torre et al. (2014, p.35) ask “does the pro-poor growth story still hold once we incorporate the missing top earners to the distribution?” Complementing survey data with information on top earners from tax data

and comparing with survey-based results for Argentina (1998-2003) and Colombia (2002-2010), the authors indeed find that inequality levels are corrected upwards. However, they also find that inequality dynamics are more prone to diverge between both scenarios during times of economic crisis than during smooth business-cycle periods. They nonetheless conclude that while “extending this exercise to the rest of the region could shed more light on the determinants of income distribution over time, we feel confident that the trends in income inequality unveiled by the household survey data are a good approximation to the real Gini for much of LAC” (De la Torre et al., 2014, p.36). While Amarante and Jiménez (2015) do not extrapolate lessons from a small sample of countries to the whole region, they recognize the similar evolution of the standard Gini and tax-adjusted Gini for the countries with survey and administrative data at their disposal (Argentina, Colombia, Uruguay). Even with the acknowledged problems of tax data (evasion, avoidance, exemptions, threshold changes), the authors think that tax data can add value to the study of inequality in the region, particularly from the perspective of top income concentration, which according to a growing number of studies has been quite persistent over time (Alvaredo, 2010; Alvaredo and Londoño-Vélez, 2013; Morgan and Souza, 2019; Burdin et al., 2020; Flores et al., 2020).¹

1.2 On the Macroeconomic Consistency of Inequality Estimates

There is a persistent gap between inequality studies and macroeconomic aggregates in the system of national accounts (SNA). The discrepancies can be seen at the level of income, wealth, and consumption, as well as in their observed growth rates (Ravallion, 2003; Deaton, 2005; Bourguignon, 2015; Nolan et al., 2019). This research has found that national income is larger and has been growing faster than the other income concepts traditionally used to study inequality. Such gaps make it hard to assess how macroeconomic growth is distributed by income groups, and to what extent existing distributional statistics (based both on surveys and tax records) are a proper representation of reality.

These gaps can attain particularly high levels in developing countries. For example, Barros et al. (2007) find a significant underestimation (27%) of household income in surveys compared to the national accounts in Brazil in 2003, but a lower underestimation of inequality (2%) after allocating the gaps via simulation. The reason for the lower impact on inequality is due to the larger share of missing transfer incomes and labor incomes, compared to missing property incomes. However, such an analysis was based on older SNA standards (the SNA-93), using an older base year (1985), in contrast to the now-available revised standards (SNA-08 and updated SNA-93). More recently, whenever survey aggregates are compared to SNA aggregates, capital incomes appear to be remarkably less covered than labor incomes (Törmälehto, 2011; Bourguignon, 2015;

¹For more details see Appendix A.1.

Flores, 2019). The extent to which these results imply an underestimation of inequality merits a re-examination.

Recent work in this field has now embarked on a process of combining data sources (surveys, national accounts, administrative registries, rich lists, etc.) through the development of two large-scale projects aiming to ensure the macroeconomic consistency of inequality estimates. On one side, following recommendations by the Canberra Group (2001) and Stiglitz et al. (2009), the Organization for Economic Co-operation and Development (OECD) started hosting periodic expert group meetings on Disparities in a National Accounts Framework (EG-DNA), focusing exclusively on the income, consumption and savings of the household sector (Fesseau and Mattonetti, 2013; Zwijsenburg et al., 2017). On the other side, the World Inequality Lab at the Paris School of Economics started publishing its own Distributional National Accounts guidelines (Alvaredo et al., 2016); alongside numerous country-case studies (see Piketty et al. (2018); Garbinti et al. (2018) for pioneering applications of the methodology and <https://wid.world> for further applications). The main difference with respect to the OECD’s approach is that DINA aims to distribute the national income of the total economy as opposed to just the household sector (for an in depth comparison of these projects see Zwijsenburg, 2019).

In Latin America there is an old tradition of aligning micro and macro data for distributional analysis, largely following the work of Altimir (1987). This seminal study critically analyzed available tax, social security and census data, as well as variety of household surveys, systematically comparing the latter with the national accounts. The author concluded that there was a 15-30% gap in aggregate household income, that could be significantly higher for certain income sources such as property income. These results were explicitly assumed to be an indicator of underestimation of each type of survey-based income. Thus, the UN-ECLAC proceed to correspondingly adjust survey-based incomes, with significant implications for inequality analysis –the Gini coefficients increased by 10-15%). Despite its positive intentions, this methodology was shown to have many caveats (Bourguignon, 2015). It was progressively abandoned by ECLAC in the last few years. The rise and fall of this experience, is the result of both the need for a reconciliation of data sources –or at least of the need to fully understand its potential consequences– and of the significant challenges of such an endeavor. Our paper seeks to address previous shortcomings with the data that are described in the following section.

2 Data sources

We rely on four main data sources: households surveys, income tax records, social security records, and national accounts. Table 1 schematically presents the data sources for countries included in this study, together with the years covered by each source, while table 2 displays data availability for countries that remain excluded from our analysis.

The following subsections elaborate on the databases used.

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 harmonization 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 labor, 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 (where post-tax is also after social contributions), the exception being Brazil, where pre-tax incomes are recorded. Part of the harmonization process involves the imputation of imputed rental income of owner-occupiers, which is absent in the surveys. This is via an internal estimation model based on matching data from similar rented dwellings.

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 harmonization processes (ECLAC-WB) are independent, they produce very similar results in terms of income distribution.³

Table 2 makes it explicit that many of the countries that remain excluded, 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).

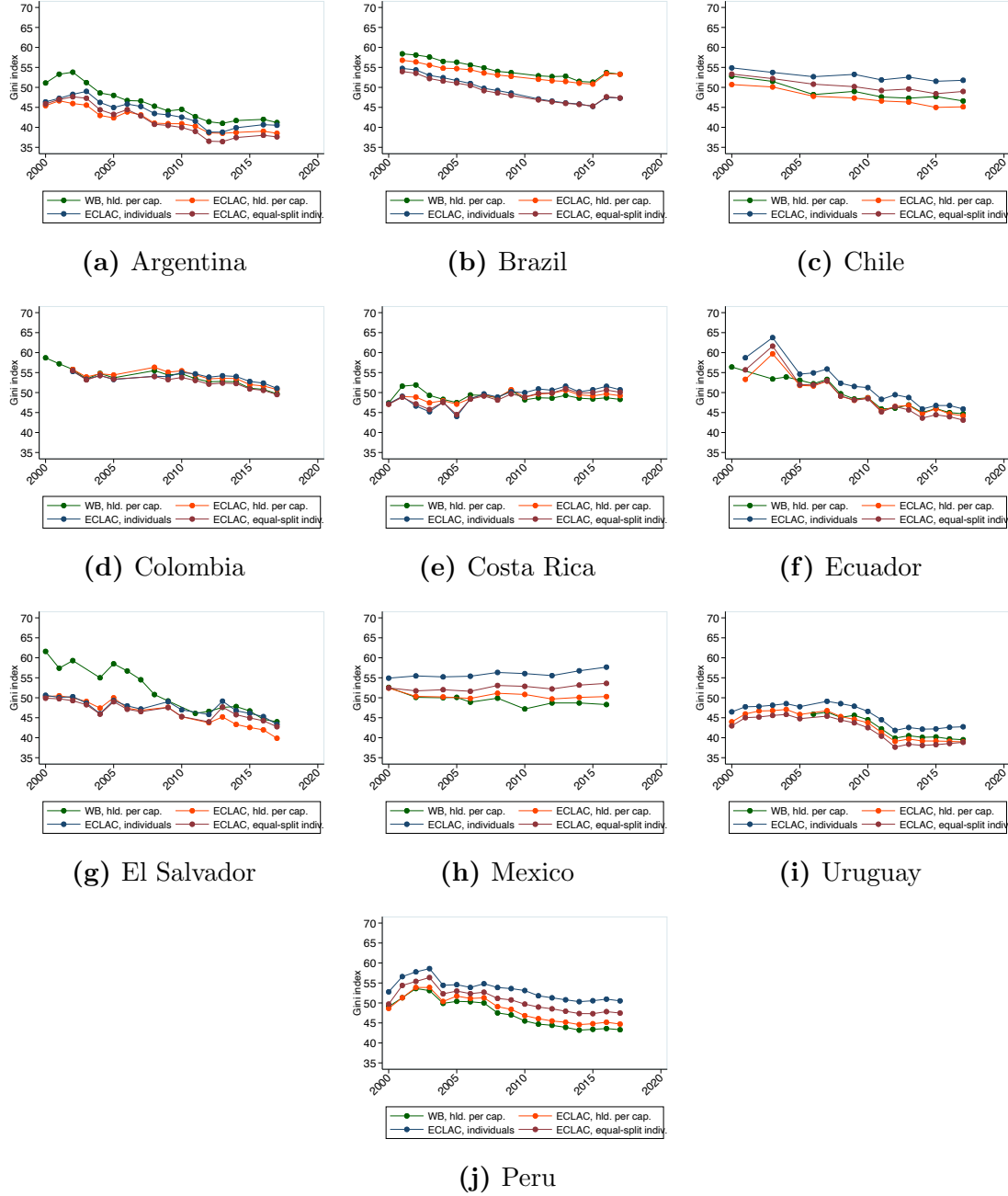
Figure 2 shows the decomposition of income in surveys, before any adjustment, in terms of wages, pensions, capital income, self-employment income, and imputed rents.

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. Both the World Bank and ECLAC use the same underlying household surveys prior to harmonization. The World Bank database is the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), produced in cooperation with the Center for Distributive, Labor and Social Studies (CEDLAS) of the Universidad Nacional de La Plata in Argentina.

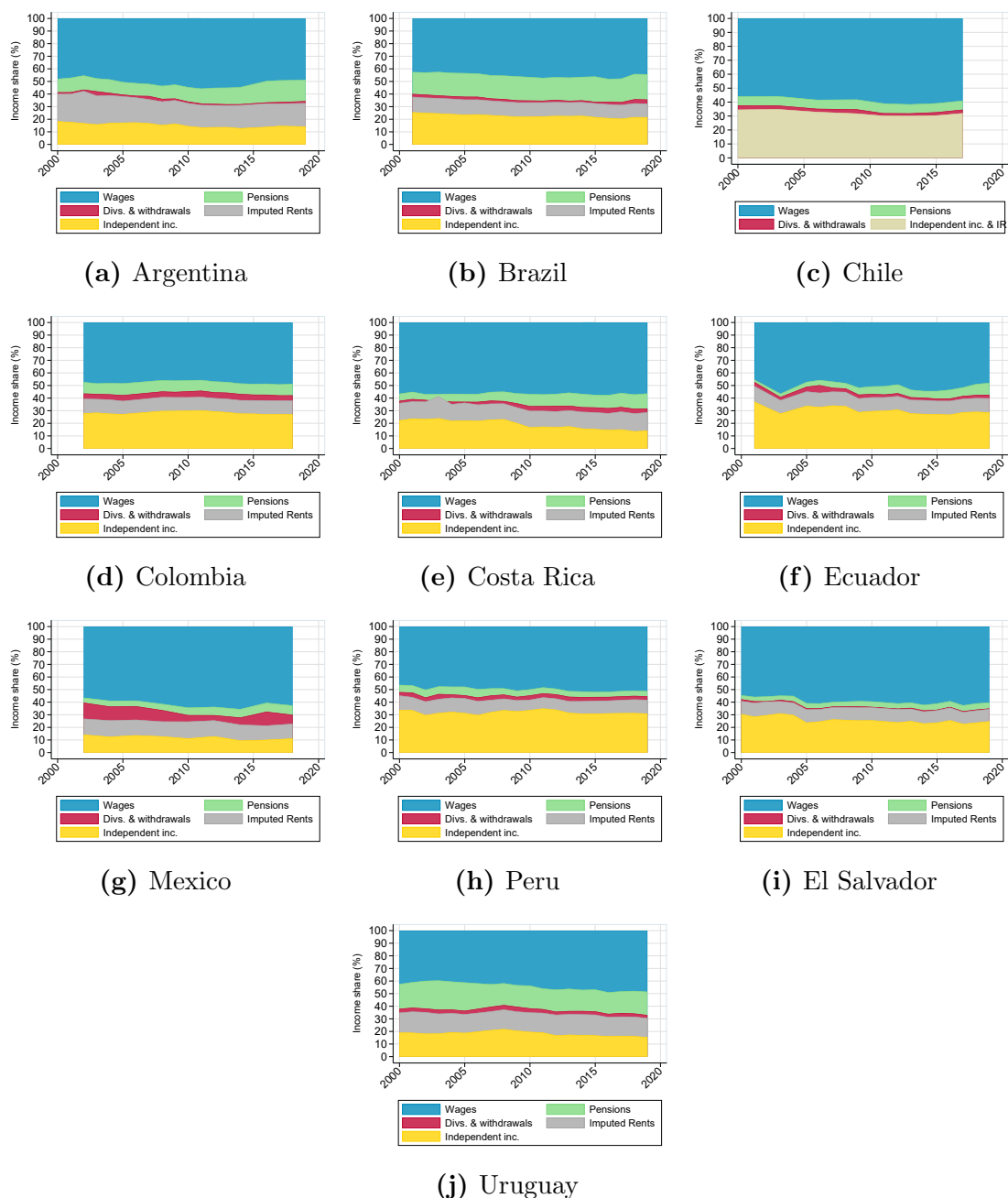
Wages and self-employment income represent 60-90% of total household incomes, while capital incomes are much lower.

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 of over 10 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).

Figure 2: Income composition - raw surveys



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)
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%
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%
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%
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%
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%
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%
Uruguay	Encuesta Continua de hogares (ECH), Instituto Nacional de Estadística (INE)	2000-2005, 2007-2019	Income tax microdata, Dirección General Impositiva	2009-2016	~75%
<i>Note.</i> Own elaboration.					

Table 2: Excluded countries

Country	Survey microdata		Availability
	Source	Sample size, thousands of individuals	
Bahamas	Bahamas Living Conditions Survey	6	2001
Belize	-	-	-
Bolivia	Encuesta de Empleo, Desempleo y Subempleo, Insituto 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), Institutio 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 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 et al., 2016), which are restricted to the top percentile of the distribution only.

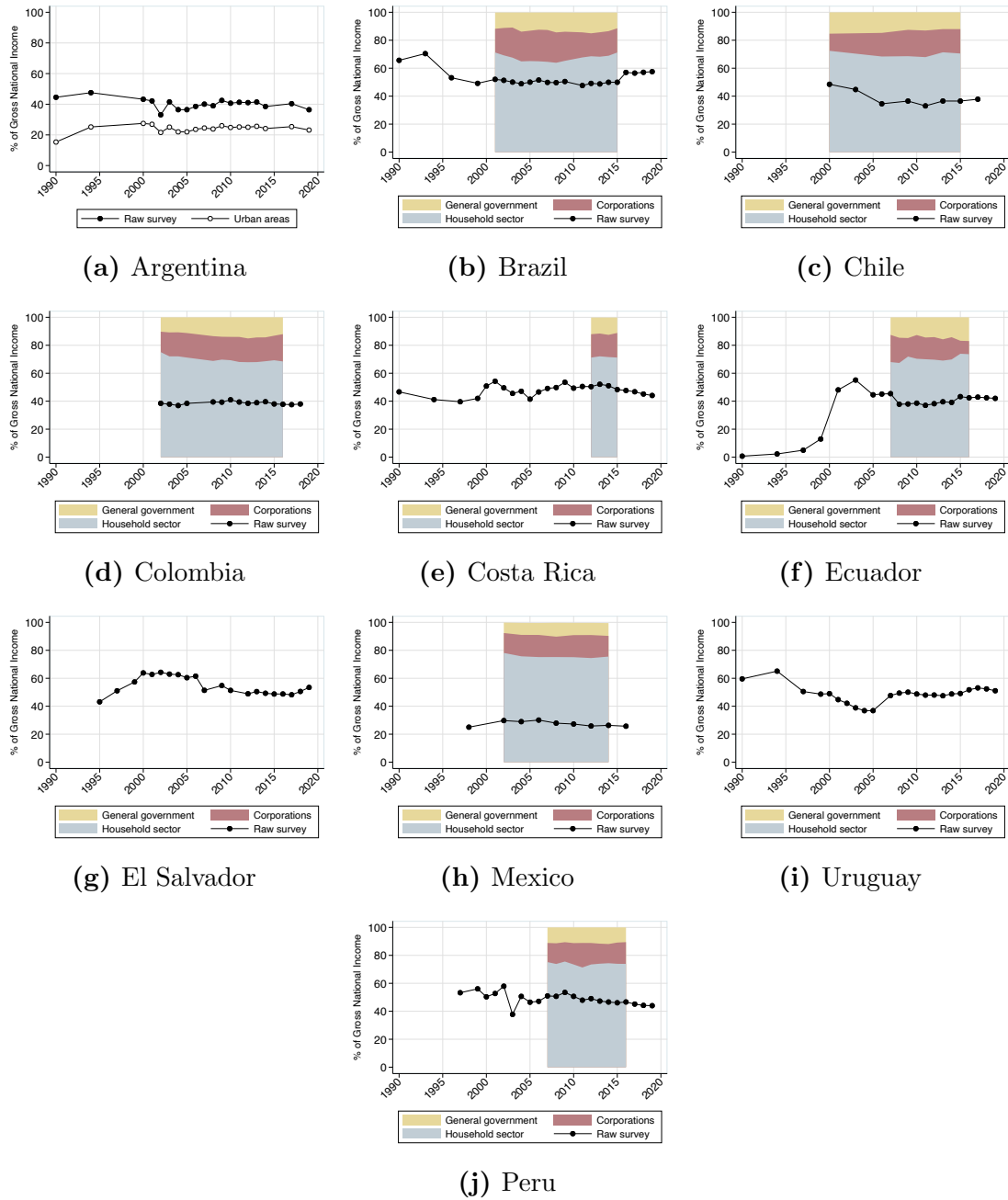
2.3 National Accounts

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 for our purpose, 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 the System of National Accounts recommends 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 property income that is imputed investment income of households from insurance, pension and collective investment 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

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.

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 adjustment, as percentage of GNI. 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 and GNI is very large, usually above 40-50%.

3 Estimation Methods

Our estimation procedure is based on four stages. We first estimate a survey-based distribution of income. The transition from this distribution to the distribution of national income as measured in the national accounts is accomplished in three subsequent 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 income components to match aggregates from the 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 these adjustment steps.⁴

3.1 Surveys adjusted with administrative data

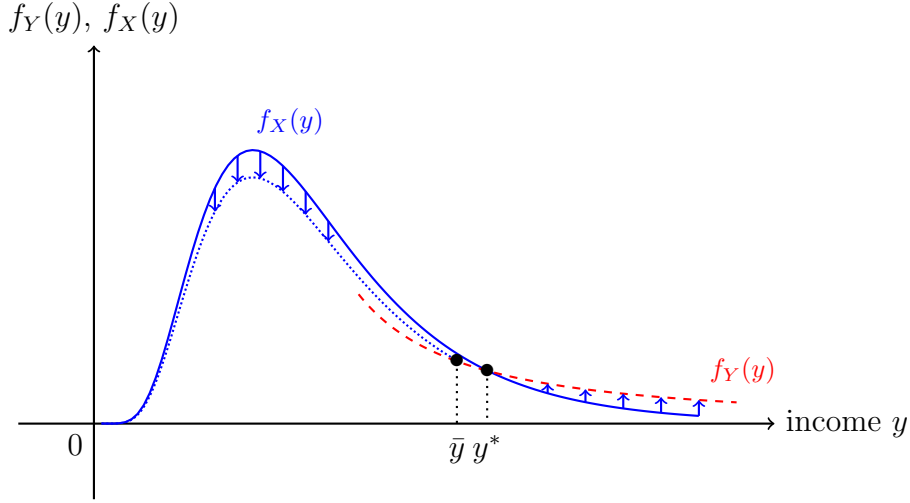
The use of administrative data refers to both personal income tax declarations and social security records. These sources are mainly used to improve the coverage of top income groups in the survey, which are often badly captured; especially when register data is not used in the surveying process, which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

In general, administrative records not only include individuals that are richer than the richest survey respondents, but also report larger numbers of moderately high 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

⁴For a more detailed description of the general procedure we employ in this paper see Alvaredo et al. (2020).

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.

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.

To 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 at the very top). For surveys where administrative records do not exist, we assume within-country stability for these coefficients to make the adjustment.

Prior to reweighting the survey we deduct tax paid from the declared income in tax data for all countries where the survey reports post-tax income (i.e. all countries except Brazil). This ensures that we are adjusting the survey using a comparable income definition.

3.2 Scaling to incomes from national accounts

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 survey incomes after adjustment with administrative data. Since the relevant macro aggregates are reported before income tax in the national accounts we add effective income tax paid across the adjusted survey distribution for the nine countries with post-tax survey incomes. Appendix A.2 explains

Table 3: Mapping household income concepts across data sets

Income in Survey [1]	National Accounts (SNA080) [2] = [3] + [4]	Matching definitions in SNA and Survey [3]	Problematic or Missing in Survey [4]
Labour	Compensation of employees (D1)	Wages, salaries* (D11)	Social security contributions (D61)
Housing rent	Operating surplus** (B2)	Rent of owner occupiers	Rental income from dwellings
Capital	Property income*** (D4)	Interests received (D41r), Dividends (D42)	Interests paid (D41u), Rent of natural resources (D45), Investment income of insurance policy holders (D441), Investment income of pension funds (D442), Investment income investment funds (D443)
Mixed	Mixed income (B3)	Self-employed income	Rent of non-dwelling buildings
Benefits	Social transfers (D62)	Pension benefits, Other cash benefits	Unemployment insurance

Listed items are pre-tax in SNA, while most of them are post-tax in surveys.

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

(**) Operating surplus is gross of depreciation.

(***) Property income is gross of property income paid (interest, rent of natural resources).

Notes: Summary table, based on United Nations (2008) and OECD (2013).

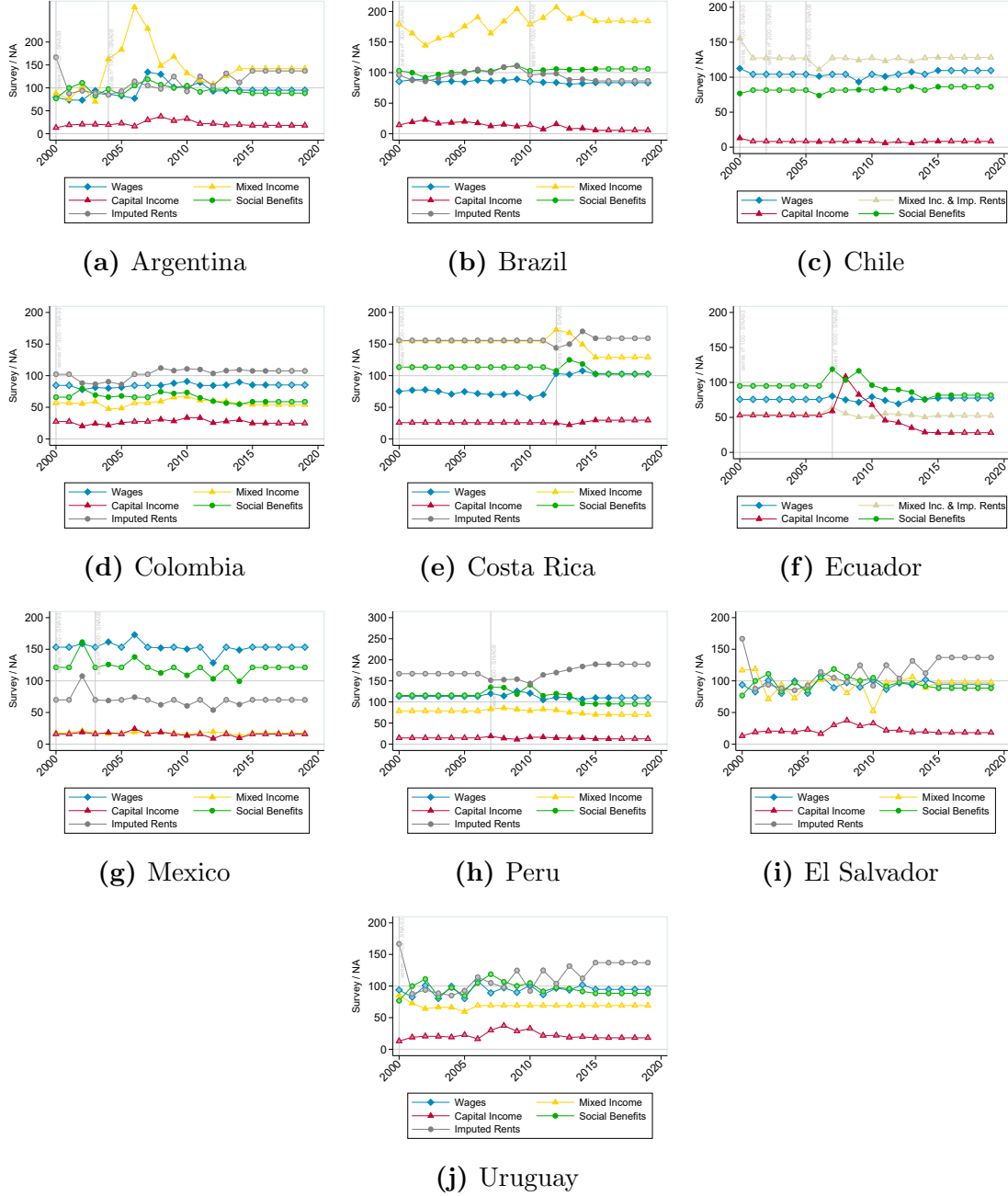
how these tax rates are computed.

Table 3 summarizes our benchmark matching of income concepts, which are gross of capital depreciation due to lack of information. For labor 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 property incomes related to investment income disbursements (D44) is not available in most countries in the region. Therefore, we scale the survey's total capital income to the property income (D4) aggregate in the national accounts, which includes incomes actually received by households (interests and dividends), as well as those received by insurance funds (D441), pension funds (D442), and investment funds (D443), which are imputed to households.⁵

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.

⁵Future estimates will attempt to disentangle these investment incomes using more detailed national accounts data for a few countries.

Figure 5: Scaling Factors for re-weighted Surveys



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 included in the net national income of the total economy. 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 households. Since other missing incomes are imputed proportionally, only retained earnings have a real distributive impact. In order to estimate this aggregate, we start from the balance of primary incomes of the corporate sector, including both the financial and non-financial sectors. We use data from WID.world given that it is more detailed than UN data. This aggregate already excludes the share of profits corresponding to portfolio investments from foreigners (which WID.world estimates (see Alvaredo et al. (2020))). 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 (D4 in the SNA). We thus subtract the same proportion from the balance of primary incomes of the corporate sector. Figure 6 displays the total amount of undistributed profits attributed to national residents, 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 our income surveys as proxies. One option would be impute them to dividends. However, too few people declare dividends in our surveys. We thus prefer to distribute them proportionally to the sum of declared dividends and employers' income, where an employer's income refers to the total income of individuals that declare being an employer in surveys 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.

Figure 6: Undistributed Profits as % of Aggregate Incomes

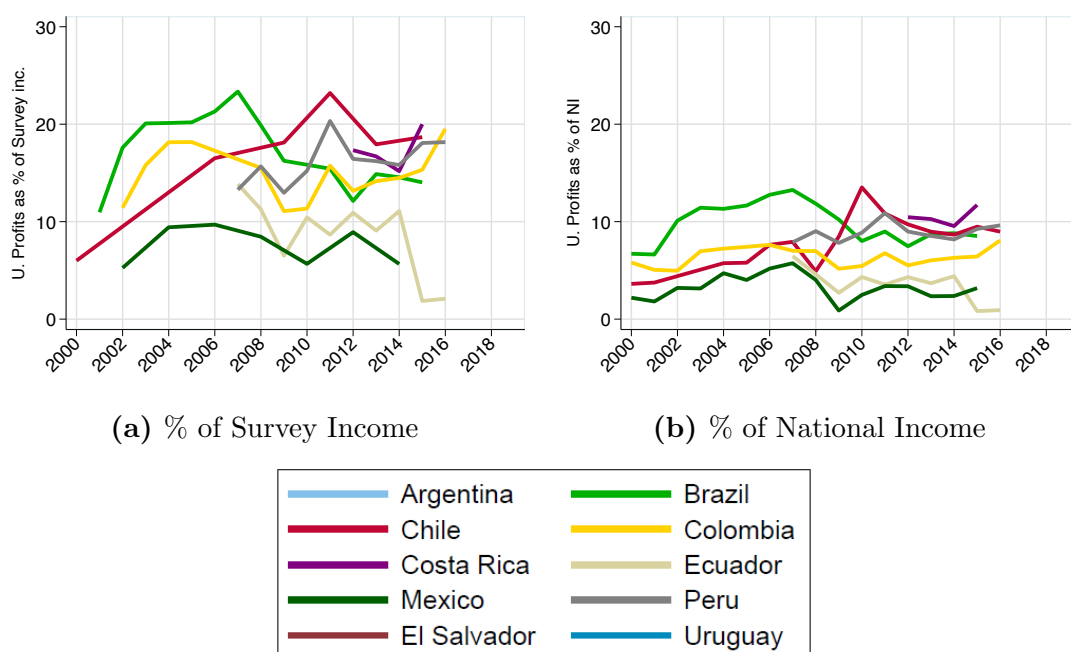
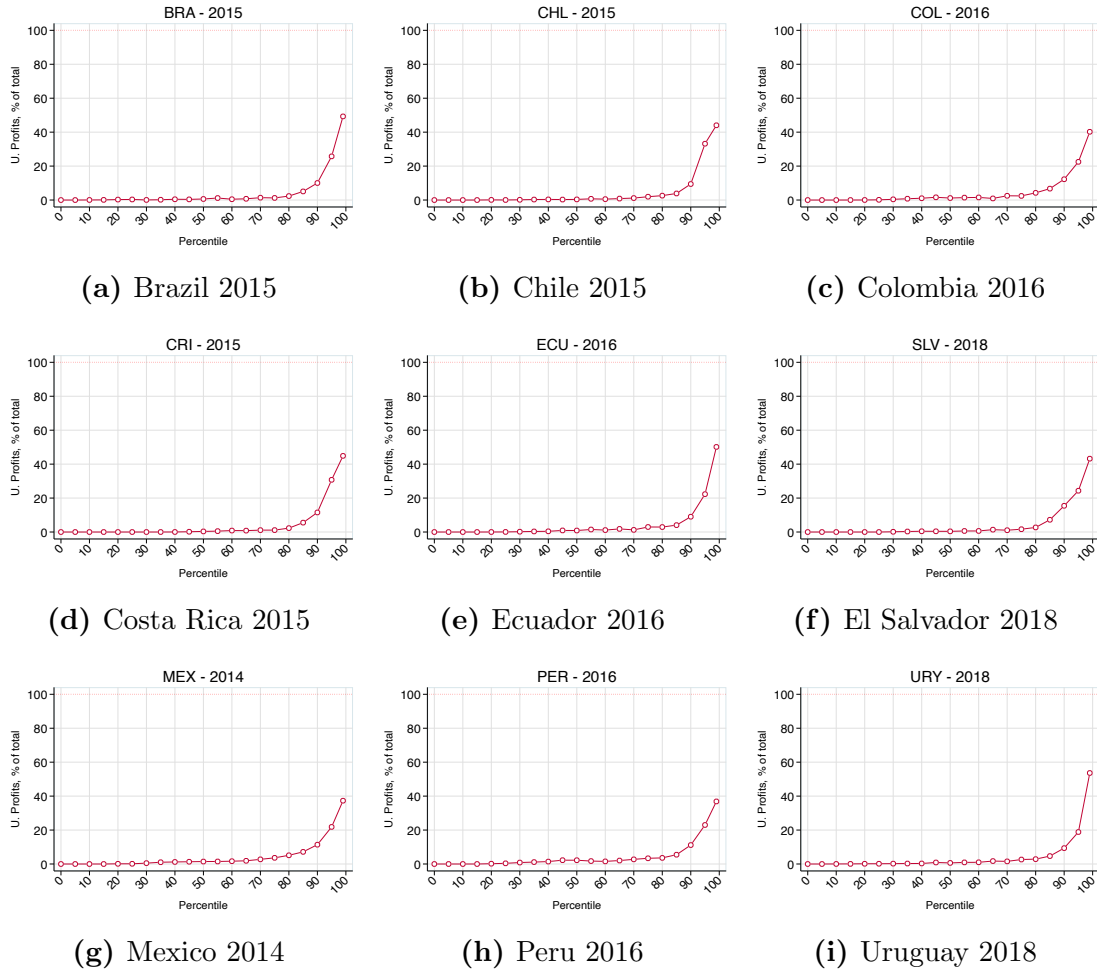


Figure 7: Share of Total Undistributed Profits Imputed to each Fractile



4 Results

4.1 A tale of four estimates

In this section we briefly present a summary of the main results. Figure 8 depicts the evolution of the Gini coefficient in our four distributions: (i) the original survey, (ii) the survey after the adjustment with administrative data, (iii) the adjusted survey scaled up to household incomes in the national accounts, (iv) and the resulting distribution after the imputation of the missing components of pre-tax national income.⁶ In all cases, our distributional estimates rank adult individuals (aged 20 and above) by increasing intervals of pre-tax income, where the total income of married couples is divided by two. Thus, our main estimates correspond to “equal-split” adult incomes (incomes of single individuals and equally-split couple incomes). Pre-tax income corresponds to wages, self-employed income, capital income, imputed rent, and pensions, net of social contributions.

Two key comments regarding the evolution of the Gini coefficients are worth highlighting. First, inequality estimates increase at each stage of the adjustment for *all* countries and years considered. The adjustment of surveys based on available tax data increases overall income inequality as a result of increasing the weight of higher income individuals. The subsequent scaling of household incomes to national accounts increases inequality, as incomes that are scaled up by higher scaling factors are precisely those that are more concentrated in the top tail (especially capital incomes, see Figure 5). The final adjustment to national income increases inequality as the result of the allocation of undistributed profits, which represent a large share of aggregate income (Figure 6), and are imputed mostly to top fractiles (Figure 7).

Second, in some cases, inequality trends after each of the three adjustment steps are consistent with the downward trends depicted by the surveys: this is clearly the case for Argentina, Colombia, Ecuador, El Salvador, Uruguay and Peru.⁷ For other countries – such as Brazil, Chile, Costa Rica, Mexico, and Peru – we observe trends that gradually flatten or even increase with each step. Furthermore, in the period after 2015 it seems that the falling inequality trend comes to a halt and even reverses in several countries.

4.2 Reconciling estimates with the literature

As discussed in Section 1, the conventional narrative is that income inequality fell in most Latin American countries during large parts of the first fifteen years of this century. To what extent do the results presented above challenge this view? To properly answer this question, we need to clarify inequality of what and among whom, and clearly distinguish

⁶Figures 9 to 12 depict income shares for each of the four distributions, which follow a similar trend.

⁷In the case of Costa Rica, the trends are also consistent with survey data but it was not downward in the first place.

between inequality levels and trends.⁸ Indeed, most estimates for Latin America (e.g. World Bank official estimates) are (1) survey-based and (2) based on per-capita household income (for the whole population). These estimates turn out to be quite different both in level and trends from an inequality series based on the distribution of national income between equal-split adults, which corresponds to our estimates after the final adjustment step. Thus, to understand the difference, we should proceed in a step-by-step comparison of the estimates.

Figure 1 showed that we are able – as expected – to consistently reproduce the downward trend in the inequality of per-capita household income using ECLAC’s harmonized surveys. For comparison, we also included individual-adult and equal-split-adult inequality series. This simple exercise shows that survey-based inequality estimates vary in levels and in many cases in trends (flatter or even reversing trends, such as in Mexico), when we move from household per-capita incomes to equal-split adult incomes.

The tax-data adjustment step is based on sources of very heterogeneous quality as commented in Section 2.2, and hence it represents a very important – considering the well documented shortcomings of surveys – but challenging step. As depicted in Figures 8 to 12, one of the key takeaways of this exercise is that most of the trend-reversals (where they exist) operate in this step, even if it does not alter the inequality levels as much as the subsequent scaling of incomes to the national accounts.

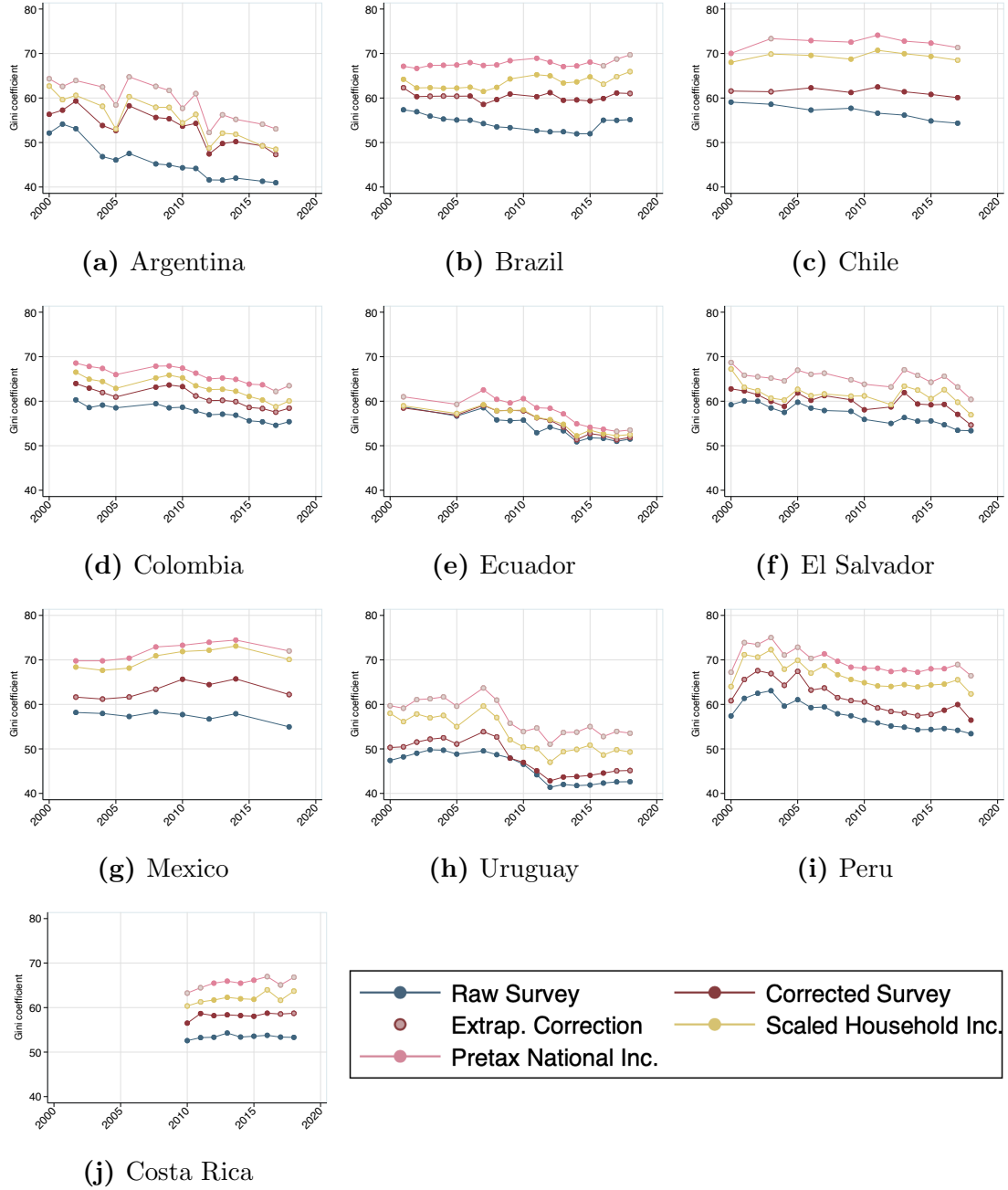
The scaling to national accounts requires a significant amount of imputation, but it is conceptually consistent with survey-based and tax-based incomes in the sense that it refers to the household sector. It is reassuring that this imputation does not have a significant impact on inequality trends, as compared to the tax-adjusted estimates (Figure 8), although it does have a substantial impact regarding levels. This implies that the mis-measurement of income in micro source is significant but constant over the time period we are assessing. Something similar can be said about the final adjustment for missing components of national income, with the important difference that the resulting distributions includes incomes that are not part of incomes accounted for by household surveys (see Section 3.3).

In general, most of the differences in the qualitative narrative about Latin American inequality is a result of (a) different units of analysis and (b) the tax-data adjustment of the upper tail in surveys. The two subsequent adjustments (to household sector income and to national income) make the estimates fully consistent with national accounts and have a strong impact on inequality levels, but have no obvious incidence on trends, which are for the most part determined in earlier stages. Differences in (a) are a matter of unit definition (each with pros and cons, shedding light on different aspects of inequality) and

⁸There is yet another important dimension of the analysis which will not be addressed in this version of the paper. It refers to the sensitivity of the results to the inequality indicator chosen. As most of the tax-based literature shows (Appendix A.1), a downturn in overall income inequality, measured for instance with the Gini coefficient, may coexist with stability or even an increase in top income shares.

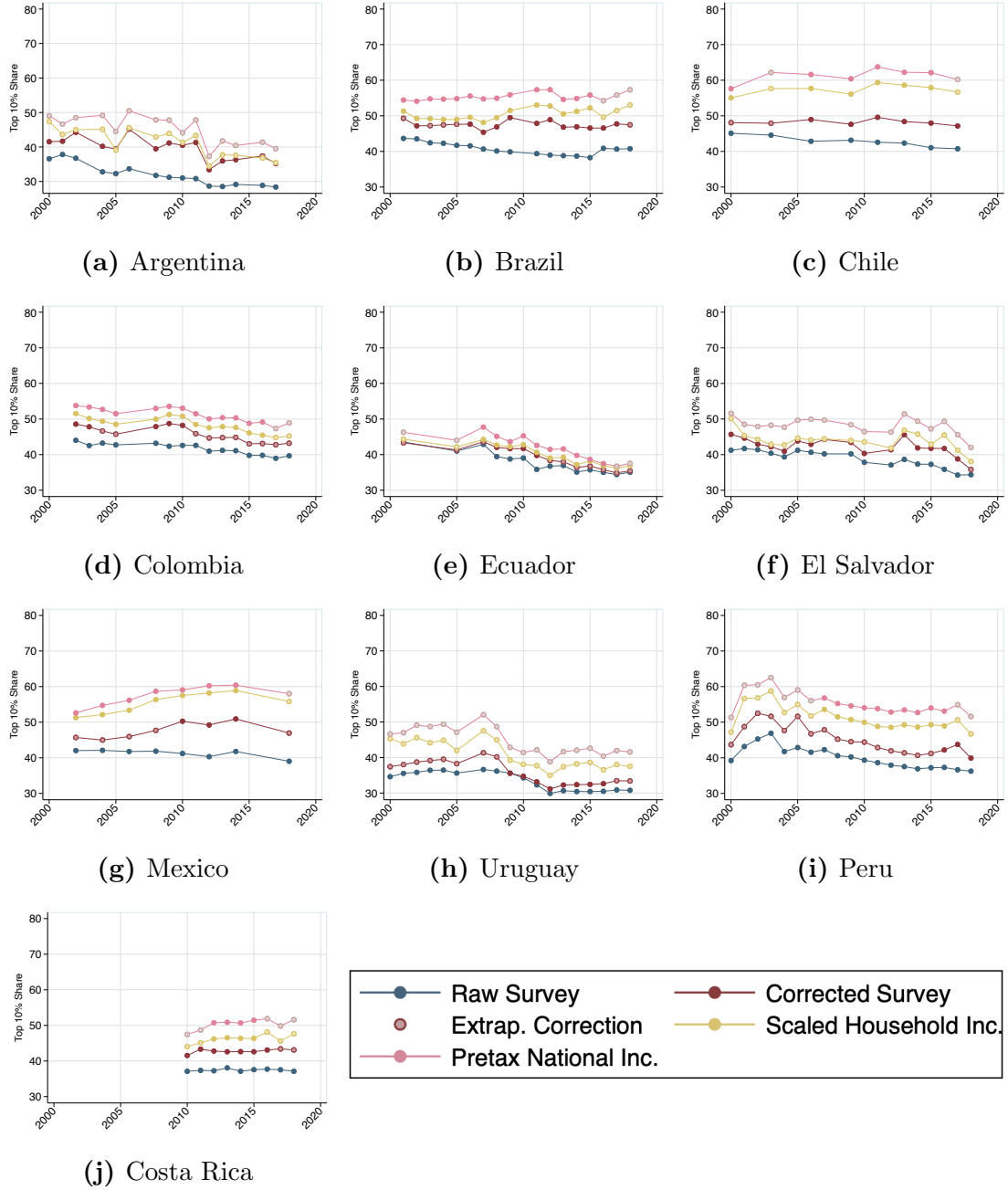
data-availability, while (b) responds to a much more decisive matter –the crucial need to appropriately account for the upper tail of the income distribution, as evidenced by the large body of literature on top incomes in Latin America (see Appendix A.1).

Figure 8: Gini coefficients in four distributions



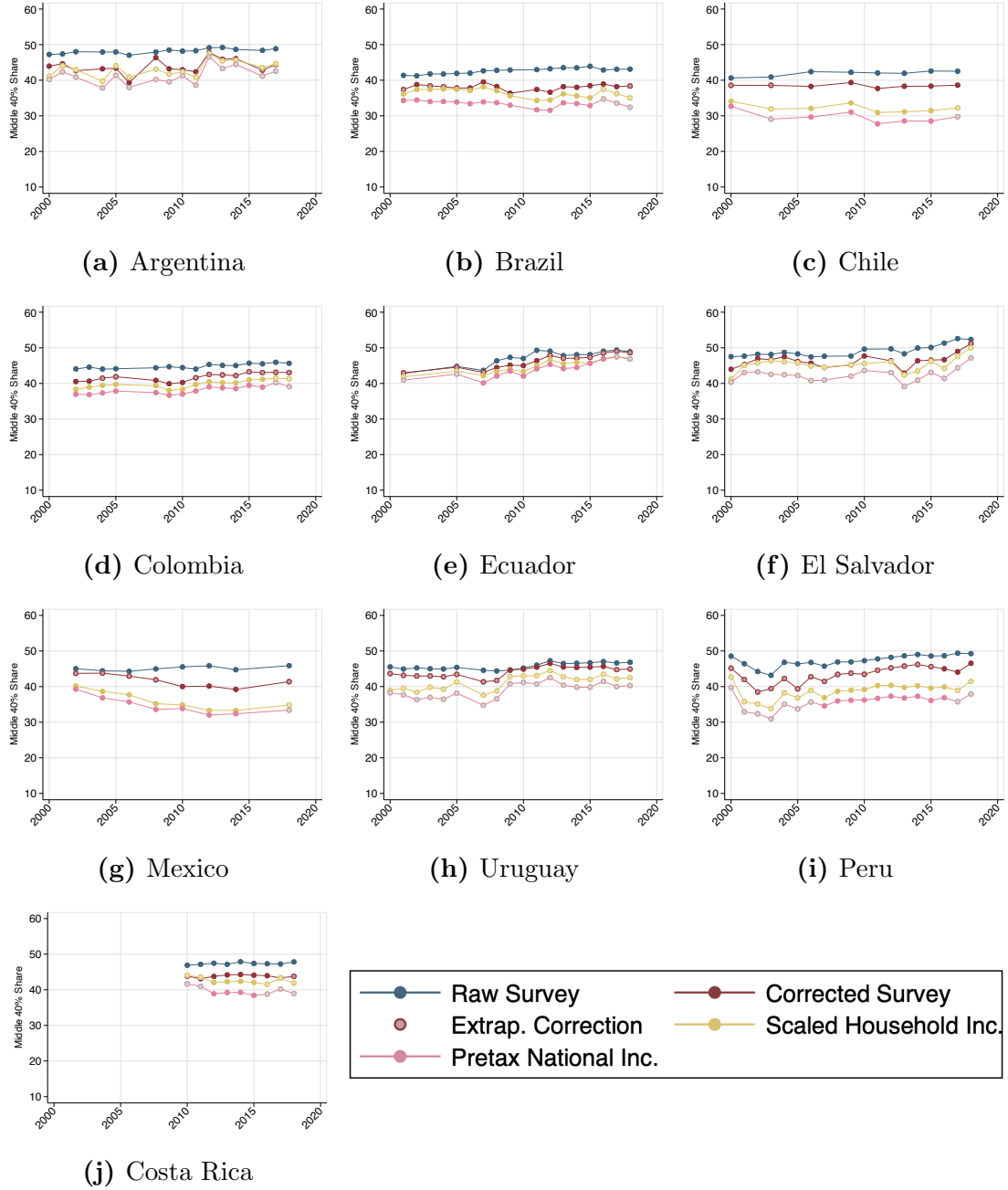
Note. Own elaboration. The figures depict four distributions: the household survey-based distribution and the three augmented distributions based on three adjustment steps to the survey. The first step uses tax data to reweight the raw survey; the second step scales the income totals in the tax-adjusted 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 adjustment step was imputed based on remaining country/year averages. Point estimates for Argentina in 2003 and 2007, Ecuador in 2003 and 2006, and Mexico in 2016 are excluded due to data inconsistencies. The distributions are of pre-tax income (including pensions and after social contributions) among individual adults (20 and over), where the total income of married couples is divided by two.

Figure 9: Top 10% Share in four distributions



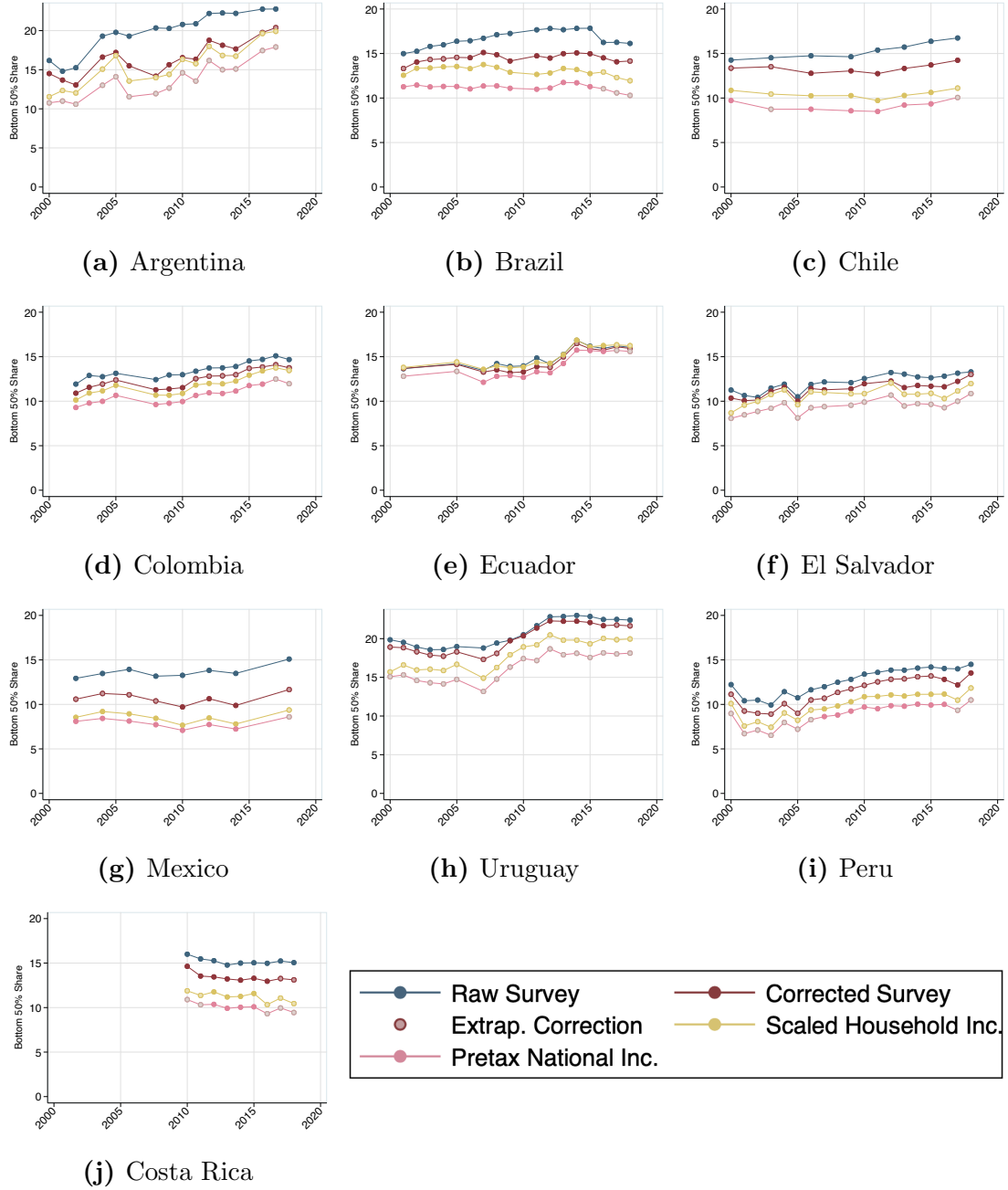
Note. Own elaboration. The figures depict four distributions: the household survey-based distribution and the three augmented distributions based on three adjustment steps to the survey. The first step uses tax data to reweight the raw survey; the second step scales the income totals in the tax-adjusted 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 adjustment step was imputed based on remaining country/year averages. Point estimates for Argentina in 2003 and 2007, Ecuador in 2003 and 2006, and Mexico in 2016 are excluded due to data inconsistencies. The distributions are of pre-tax income (including pensions and after social contributions) among individual adults (20 and over), where the total income of married couples is divided by two.

Figure 10: Middle 40% Share in four distributions



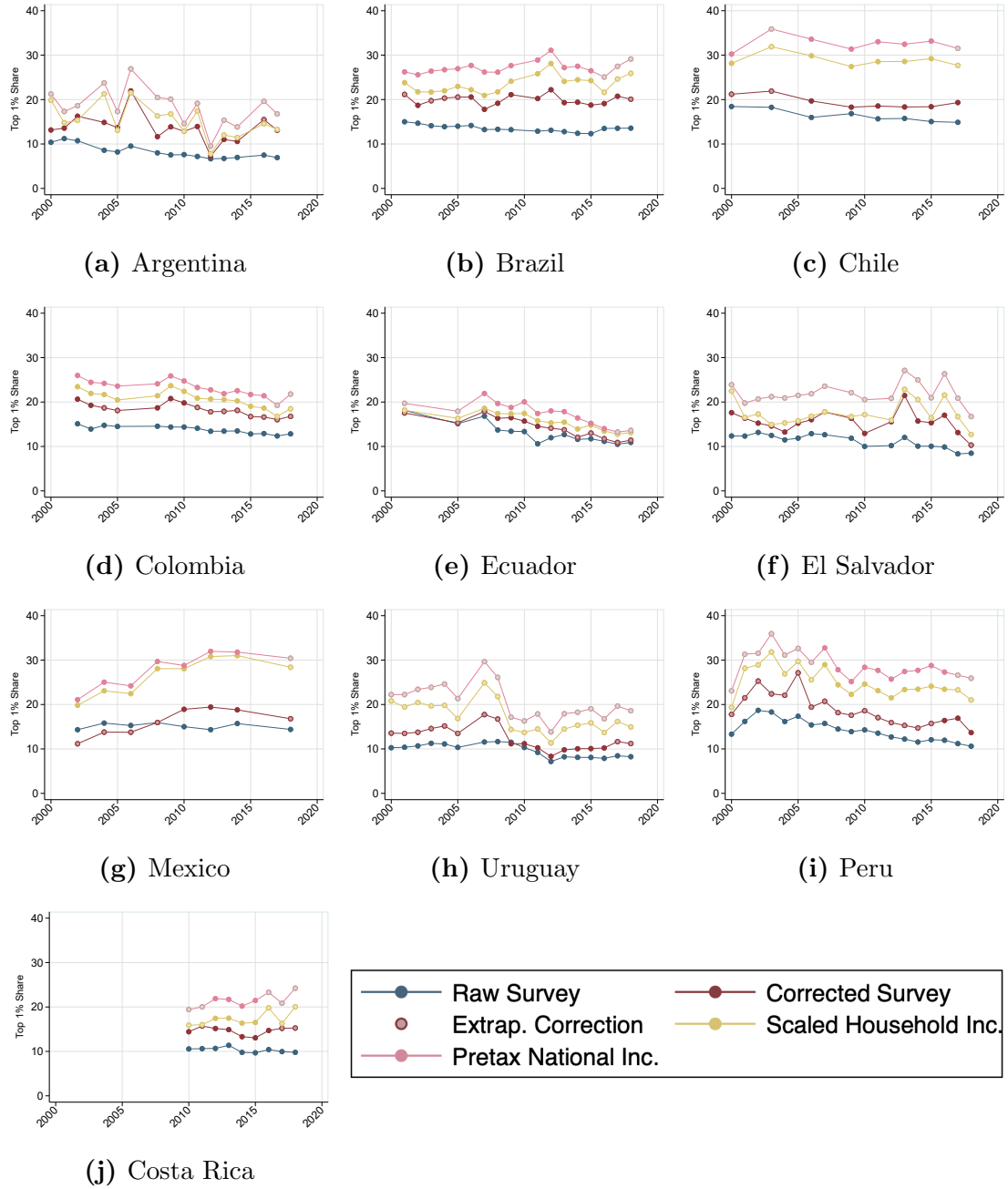
Note. Own elaboration. The figures depict four distributions: the household survey-based distribution and the three augmented distributions based on three adjustment steps to the survey. The first step uses tax data to reweight the raw survey; the second step scales the income totals in the tax-adjusted 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 adjustment step was imputed based on remaining country/year averages. Point estimates for Argentina in 2003 and 2007, Ecuador in 2003 and 2006, and Mexico in 2016 are excluded due to data inconsistencies. The distributions are of pre-tax income (including pensions and after social contributions) among individual adults (20 and over), where the total income of married couples is divided by two.

Figure 11: Bottom 50% Share in four distributions



Note. Own elaboration. The figures depict four distributions: the household survey-based distribution and the three augmented distributions based on three adjustment steps to the survey. The first step uses tax data to reweight the raw survey; the second step scales the income totals in the tax-adjusted 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 adjustment step was imputed based on remaining country/year averages. Point estimates for Argentina in 2003 and 2007, Ecuador in 2003 and 2006, and Mexico in 2016 are excluded due to data inconsistencies. The distributions are of pre-tax income (including pensions and after social contributions) among individual adults (20 and over), where the total income of married couples is divided by two.

Figure 12: Top 1% Share in four distributions



Note. Own elaboration. The figures depict four distributions: the household survey-based distribution and the three augmented distributions based on three adjustment steps to the survey. The first step uses tax data to reweight the raw survey; the second step scales the income totals in the tax-adjusted 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 adjustment step was imputed based on remaining country/year averages. Point estimates for Argentina in 2003 and 2007, Ecuador in 2003 and 2006, and Mexico in 2016 are excluded due to data inconsistencies. The distributions are of pre-tax income (including pensions and after social contributions) among individual adults (20 and over), where the total income of married couples is divided by two.

5 Concluding remarks

Having reached this point, we should now be in a position to answer the question posed at the beginning: is Latin America more unequal than we thought or is it not as rich? It is worth stressing once again that this exercise relies on imperfect and heterogeneous data alongside numerous assumptions. However, it is also true that it represents a unique attempt to make use of all available data sources in a coherent manner to provide conceptually consistent inequality estimates. Moreover, it is an honest effort to build a bridge between different inequality approaches. In this sense, it should be regarded as a contribution to open a debate on an important topic and not to close it.

Regarding inequality levels, it is clear that only one of the following is true: either Latin America is more unequal than we thought it was, or it is not as rich. The large gap between the second and third distributions we estimate, i.e. after adjusting the survey using tax data and scaling to the household sector aggregates in the national accounts, shows that the seminal findings by Altimir (1987) are still essentially true. Consistent with these earlier results, we can say is that there is a large gap between micro and macro incomes that cannot be fully explained by conceptual differences. However, unlike what was believed before, we cannot claim that measurement error is exclusively a micro-data problem – it may also affect macro-data. An additional levels-gap emerges when all remaining incomes, especially undistributed profits, are imputed in the last stage. The credibility of this final adjustment will depend on the reader’s belief on the importance of imputing this income to suspected firm owners. From our perspective, regardless of the accounting convention on whether to allocate this income to firms or to their owners (which may largely be affected by tax incentives and other factors), it is evident that they are resources controlled by individuals and they should be accounted for in any meaningful inequality analysis, if only to avoid cross-country biases affecting the distribution of profits.

From the perspective of inequality levels, the question therefore is still open, and the answers are mutually exclusive in our opinion. On the contrary, when trends are factored in, our findings suggests that the original question may be answered as ‘both’. We find trends increasing or at least flattening for half of our sample, and also – no less important – inequality falling in a fully consistent manner with the original surveys, which do not seem to be so off the mark. At the same time, most of the changes we see are already visible in the tax-adjustment stage of the exercise, i.e. before even considering a single macro aggregate from the national accounts. Therefore, we can argue that inequality in Latin America did not fall in all countries as we thought it did, but this seems to be the case independently of making the estimates macro-consistent.

Naturally, the above conclusions are highly dependent on the particular assumptions made. We think that they can be strongly defended, but we do recognize that others can be made. Considering that surveys miss half of national income, we are perfectly

aware that many other distributions can theoretically be estimated with a different set of assumptions. Further research is still needed at the country-level –exploiting the rich country data lost in our generalized approach and the local knowledge of researchers– to provide a full picture of what was probably one of the most significant economic episodes in Latin American history.

A Appendix

A.1 Literature on top incomes using administrative data

A.1.1 Argentina

Alvaredo (2010), covering the period 1932-2004, is the seminal reference on the topic, with no precedent to our knowledge. This line of work was recently picked up again by Jiménez and Rossignolo (2019), who similarly use tax registries alongside updated national accounts statistics, for the period 2004-2015. The latter emphasize certain caveats regarding the use of statistical information, which they deem to be “scarce, incomplete, inconsistent or still nonexistent.”

A.1.2 Brazil

Mortara (1949) was the first scholar to use personal income tax records in Brazil, applying the Pareto interpolation to tabulated data to study income inequality. His contribution did not spur further studies until the 1970s, when scholars with ties to the military dictatorship, such as Kingston and Kingston (1972) and Langoni (1973), also relied on income tax data to try to push for more benign views of the rise in inequality in the 1960s. The use of tax records to study top incomes would not re-surface until the 2010s when newly-released income tax tabulations became available to researchers. Not only did this data show that surveys exaggerated the fall in inequality in the 2000s (Medeiros et al., 2015; Morgan, 2017), it was also used to measure distributional effects of taxation (e.g. Castro and Bugarin (2017); Gobetti and Orair (2017); Fernandes et al. (2018)). Coupled with archival data on historical income tax tabulations, this new data was used by Souza and Medeiros (2015), Morgan (2015) and Souza (2016, 2018) to estimate top income shares in the long-run for the first time. While the combination of survey and tax data into a single measure of inequality was attempted by Medeiros et al. (2015); Souza (2016); Medeiros et al. (2018), their reconciliation with national income statistics over the 2000s was studied by Morgan (2017) and by Morgan and Souza (2019) over the long run.

A.1.3 Chile

The earliest attempt to study top income trends did not come from the use of administrative tax data but from surveys (Sanhueza and Mayer, 2011). López et al. (2013) were the first scholars to employ personal income tax tabulations to study top incomes over the 2000s. Administrative microdata of tax declarations were used by ? to better study top incomes in the context of an institutional set-up tailored for the retention of a large amount of corporate profits not included in income tax returns for two individual years, refining the similar estimates made by López et al. (2013). Flores, Sanhueza, Atria, and Mayer (Flores

et al.) has been to date the most comprehensive study on top incomes, combining features from previous attempts – long run estimates from income tax tabulations (1964-2017) with imputations of retained earnings from national accounts.

A.1.4 Colombia

Londoño-Vélez (2012) was the first work to incorporate income tax databases, which were used in Alvaredo and Londoño-Vélez (2013) for the study of top incomes and their composition between 1993 and 2010. The latter reconciled the results with survey-based measures using Gini-adjustment methods from Atkinson (2007) and Alvaredo (2011).

A.1.5 Costa Rica

Zuñiga-Cordero (2018) is the first study to use multiple administrative sources of income (social security records, income tax data, national accounts) to study inequality, alongside household surveys, for Costa Rica, for the 2000-2017 period.

A.1.6 Ecuador

Few studies exist for the analysis of top incomes, with Cano (2015) initiating the trend based on microdata from tax registries over the period 2004-2010. This attempt was followed by Rossignolo et al. (2016), updating the previous series to 2014.

A.1.7 Mexico

Alvaredo et al. (2017) is the only study that used income tax data on universe of personal income taxpayers from the Mexican Tax Administration Service (SAT) and formal wage data from the universe of employer-reported information in the Declaración Informativa Múltiple (DIM) from 2009 to 2014. The authors perform a comparative analysis of incomes declared in these administrative datasets with those reported in the household surveys (ENIGH) for the same years in order to explore a potential reconciliation.

A.1.8 Uruguay

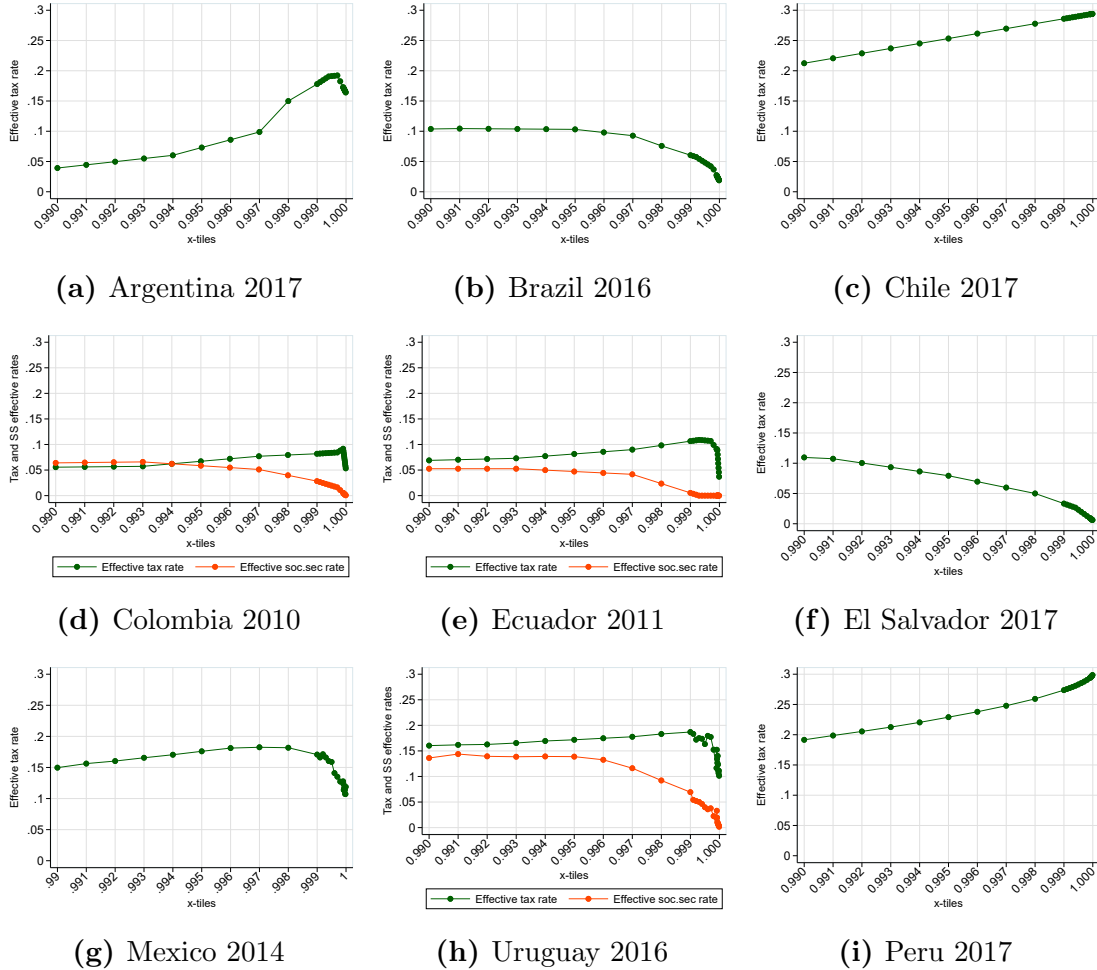
The decrease in income inequality shown in household surveys (e.g. Cornia (2014)) has been confirmed by the use of income tax records (Burdín et al., 2014; Burdin et al., 2020). Falling inequality also emerged from Distributional National Accounts (DINA) estimations (De Rosa and Vilá, 2020).

A.2 Estimation of pre-tax incomes in surveys

The inequality estimates we present in this paper concern pre-tax incomes. However, the main data source on which our estimates are based are harmonized household surveys,

which account for post-tax incomes in Latin America.⁹ In order to scale incomes to their pre-tax aggregates in the national accounts it is necessary to calculate pre-tax incomes in surveys.

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



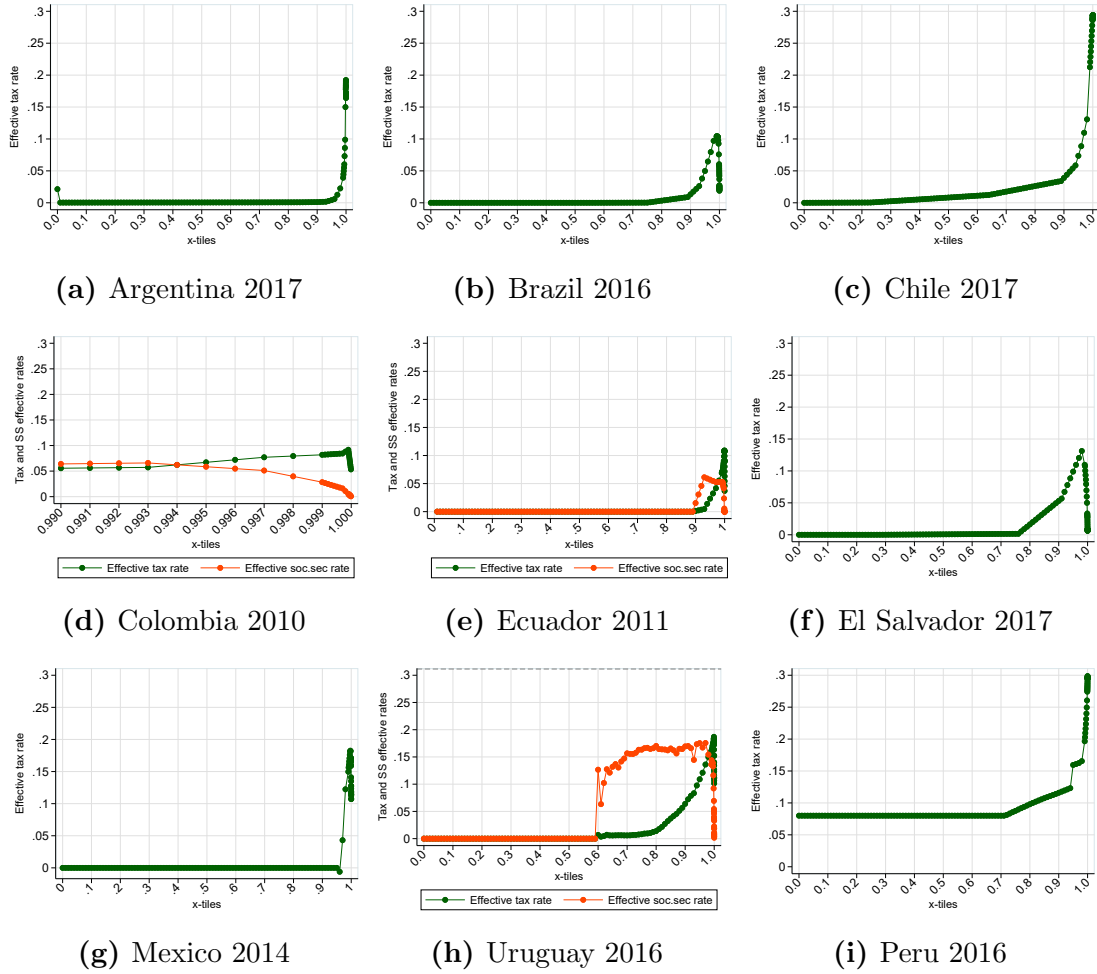
As data on direct taxes paid by households is not collected in surveys we tax data to estimate pre-tax incomes. Broadly speaking, we compute effective tax rates by income fractile in the tax data, and use these tax rates to calculate pre-tax incomes in the survey, based on the income fractiles to which individuals belong to.¹⁰ Effective tax rates by income fractile are computed for the years for which we have access to income tax data, and the average effective tax rate by fractile is used to calculate pre-tax incomes when this data is not available.¹¹ Tax data quality and coverage, however, varies significantly across

⁹The only exception is Brazil, whose survey accounts for pre-tax incomes.

¹⁰We 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.

¹¹This 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.

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



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 shown.

In the cases where 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 the same studies we use to extract top income information – Londoño-Vélez (2012) for Colombia or Cano (2015); Rossignolo et al. (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 very right tail of the distribution for most countries. In countries where this is not the case (e.g. Argentina and Chile), 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

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.

for contributions). The absence of information on social contributions is not problematic, given that the income definition we use in our estimates includes social security transfers, net of social contributions.

References

- Altimir, O. (1987). Income distribution statistics in latin america and their reliability. *Review of Income and Wealth* 33(2), 111–155.
- Alvaredo, F. (2010). The rich in Argentina over the Twentieth Century, 1932-2004. In A. B. Atkinson and T. Piketty (Eds.), *Top incomes: A global perspective*, pp. 253–298. Oxford University Press.
- Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters* 110(3), 274–277.
- Alvaredo, F., A. B. Atkinson, T. Blanchet, L. Chancel, L. Bauluz, M. Fisher-Post, I. Flores, B. Garbinti, J. Goupille-Lebret, M.-T. Clara, M. Morgan, T. Neef, T. Piketty, A.-S. Robilliard, E. Saez, L. Yang, and G. Zucman (2020). Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database.
- Alvaredo, F., A. B. Atkinson, L. Chancel, T. Piketty, E. Saez, and G. Zucman (2016). Distributional National Accounts (DINA) Guidelines : Concepts and Methods used in WID.world.
- Alvaredo, F., S. Garriga, and F. Pinto (2017). Household Surveys, Administrative Records, and National Accounts in Mexico 2009–2014. Is a Reconciliation Possible?
- Alvaredo, F. and J. Londoño-Vélez (2013). High Incomes and Personal Taxation in a Developing Economy: Colombia (1993-2010).
- Amarante, V. and J. P. Jiménez (2015). Desigualdad, concentración y rentas altas en américa latina. In J. P. Jiménez (Ed.), *Desigualdad, concentración del ingreso y tributación sobre las altas rentas en América Latina*, pp. 13–47. Santiago: CEPAL.
- Atkinson, A. B. (2007). Measuring Top Incomes: Methodological Issues. In *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*, pp. 18–42.
- Barros, R. P. d., S. Cury, and G. Ulyssea (2007). A desigualdade de renda no brasil encontra-se subestimada? uma análise comparativa usando pnad, pof e contas nacionais. In R. P. d. Barros, M. N. Foguel, and G. Ulyssea (Eds.), *Desigualdade de Renda no Brasil: Uma Análise da Queda Recente*, Volume 1, pp. 237–273. Brasilia: IPEA.
- Bértola, L. and J. Williamson (2017). *Has Latin American inequality changed direction?: Looking over the long run*.
- Blanchet, T., I. Flores, and M. Morgan (2018). The weight of the rich: Improving surveys using tax data. *WID.world Working Paper Series 2018/12* (October).

- Bourguignon, F. (2015, 12). Appraising income inequality databases in Latin America. *The Journal of Economic Inequality* 13(4), 557–578.
- Burdín, G., F. Esponda, A. Vigorito, F. Alvaredo, V. Amarante, M. Bucheli, J. Carlos Gómez Sabaini, J. Pablo Jiménez, N. Lustig, J. Onrubia, N. Noya, D. Rossignolo, and G. Pacheco (2014). Inequality and Top Incomes in Uruguay: A Comparison Between Household Surveys and Income Tax Micro-Data.
- Burdin, G., M. D. Rosa, A. Vigorito, and J. Vilá (2020). Was Falling Inequality in All Latin American Countries a Data-Driven Illusion? Income Distribution and Mobility Patterns in Uruguay 2009-2016. IZA Discussion Papers 13070, Bonn.
- Canberra Group (2001). *Expert Group on Household Income Statistics, The Canberra Group : final report and recommendations*.
- Cano, L. (2015). Income mobility in Ecuador new evidence from individual income tax returns.
- Castro, F. A. and M. Bugarin (2017). A progressividade do imposto de renda da pessoa física no brasil. *Estudos Econômicos* 47(2), 259–293.
- Cornia, G. A. (2014). Inequality trends and their determinants: Latin America over 1990 - 2010.
- De la Torre, A., E. L. Yeyati, G. Beylis, T. Didier, C. R. Castelan, and S. Schmukler (2014, 10). Inequality in a Lower Growth Latin America. Technical report, World Bank.
- De Rosa, M. and J. Vilá (2020). Distributing the missing third: growth and falling inequality in uruguay 2009-2016. *Serie Documentos de Trabajo; 05/20*.
- Deaton, A. (2005, 2). Measuring Poverty in a Growing World (or Measuring Growth in a Poor World). *Review of Economics and Statistics* 87(1), 1–19.
- Fernandes, R. C., B. P. C. Diniz, and F. G. Silveira (2018). The distributive impact of income taxes in brazil. Technical report, International Policy Centre for Inclusive Growth (IPC-IG), Working Paper, No. 171.
- Fesseau, M. and M. L. Mattonetti (2013). Distributional measures across household groups in a national accounts framework.
- Flores, I. (2019). Measuring Capital-Labour Shares and Inequality: Increasing Gaps between National Accounts and Micro-data.
- Flores, I., C. Sanhueza, J. Atria, and R. Mayer. Top Incomes in Chile: A Historical Perspective on Income Inequality, 1964–2017. *Review of Income and Wealth*.

- Flores, I., C. Sanhueza, J. Atria, and R. Mayer (2020). Top Incomes in Chile: A Historical Perspective on Income Inequality, 1964–2017. *Review of Income and Wealth* 66(4), 850–874.
- Garbinti, B., J. Goupille-Lebret, and T. Piketty (2018, 6). Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA). *Journal of Public Economics* 162, 63–77.
- Gasparini, L., J. Bracco, L. Galeano, and M. Pistorio (2018). Desigualdad en países en desarrollo: ¿ajustando las expectativas? *Documentos de Trabajo del CEDLAS*.
- Gobetti, S. and R. Orair (2017). Taxation and distribution of income in brazil: new evidence from personal income tax data. *Revista de Economia Política* 37(2), 267–286.
- Jiménez, J. P. and D. Rossignolo (2019, 1). Concentración del ingreso y desigualdad según registros tributarios en américa latina: el caso de la argentina. Technical report, Centro de Estudios para el Cambio Estructural (CECE), Buenos Aires.
- Kingston, J. and L. Kingston (1972). A distribuição da renda no brasil, 1960-1970. *Revista Brasileira de Economia* 26(4), 241–256.
- Langoni, C. (1973). *Distribuição da renda e desenvolvimento economico do Brasil*. Rio de Janeiro: Editora Expressao e Cultura.
- Londoño-Vélez, J. (2012). Income and wealth at the top in colombia: An exploration of tax records 1993-2010. Master’s dissertation, Paris School of Economics, Paris.
- López, R., E. Figueroa, and P. Gutiérrez (2013). La Parte del León: Nuevas Estimaciones de la Participación de los Súper Ricos en el Ingreso de Chile.
- López-Calva, L. F. and N. C. Lustig (Eds.) (2010). *Declining Inequality in Latin America: A Decade of Progress?* Brookings Institution Press and United Nations Development Programme.
- Medeiros, M., J. de Castro Galvão, and L. de Azevedo Nazareno (2018). Correcting the Underestimation of Top Incomes: Combining Data from Income Tax Reports and the Brazilian 2010 Census. *Social Indicators Research* 135(1), 233–244.
- Medeiros, M., P. Souza, and F. A. d. Castro (2015). O topo da distribuição de renda no brasil: Primeiras estimativas com dados tributarios e comparação com pesquisas domiciliares (2006-2012). *Dados* 58(1), 7–36.
- Messina, J. and J. Silva (2017). *Wage Inequality in Latin America: Understanding the Past to Prepare for the Future*. The World Bank.

- Morgan, M. (2015). Income concentration in a context of late development: an investigation of top incomes in brazil using tax records, 1933-2013. Master's dissertation, Paris School of Economics, Paris.
- Morgan, M. (2017). Desigualdade de Renda, Crescimento e Tributação da Elite no Brasil: Novas Evidências Reunindo Dados de Pesquisas Domiciliares e Dados Fiscais. In J. R. Afonso, R. O. Orair, and F. G. Silveira (Eds.), *Tributação e Desigualdade*. Rio de Janeiro: Letramento FGV Direito Rio.
- Morgan, M. and P. Souza (2019). Distributing growth during late-development: New evidence on long-run inequality in brazil.
- Mortara, G. (1949). Representação analítica das distribuições dos contribuintes e das respectivas rendas líquidas determinadas para a aplicação do imposto de renda, em função do valor da renda líquida. *Revista Brasileira de Economia* 3(2), 7–34.
- Nolan, B., M. Roser, and S. Thewissen (2019, 9). GDP Per Capita Versus Median Household Income: What Gives Rise to the Divergence Over Time and how does this Vary Across OECD Countries? *Review of Income and Wealth* 65(3), 465–494.
- OECD (2013). *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*. OECD Publishing.
- Piketty, T., E. Saez, and G. Zucman (2018, 5). Distributional National Accounts: Methods and Estimates for the United States*. *The Quarterly Journal of Economics* 133(2), 553–609.
- Ravallion, M. (2003, 8). Measuring Aggregate Welfare in Developing Countries: How Well Do National Accounts and Surveys Agree? *Review of Economics and Statistics* 85(3), 645–652.
- Rodríguez-Castelán, C., L. F. López-Calva, N. Lustig, and D. Valderrama (2016). Understanding the dynamics of labor income inequality in latin america. Technical report, Policy Research Working Paper 7795, The World Bank.
- Rossignolo, D., N. Oliva, and N. Villacreses (2016). Cálculo de la concentración de los altos ingresos sobre la base de los datos impositivos: un análisis para el Ecuador.
- Sanhueza, C. and R. Mayer (2011). Top Incomes in Chile using 50 years of household surveys: 1957-2007. *Estudios de economía* 38(1), 169–193.
- Souza, P. (2016). *A desigualdade vista do topo: a concentração de renda entre os ricos no Brasil, 1926–2013*. Phd dissertation in sociology, Universidade de Brasília, Brasília.

- Souza, P. (2018). A history of inequality: top incomes in brazil, 1926–2015. *Research in Social Stratification and Mobility* 57, 35–45.
- Souza, P. and M. Medeiros (2015). Top income shares and inequality in Brazil, 1928–2012. *Sociologies in Dialogue* 1(1), 119–132.
- Stiglitz, J., A. K. Sen, and J.-P. Fitoussi (2009). The measurement of economic performance and social progress revisited: reflections and overview. Technical report.
- Törmälehto, V.-M. (2011). LIS and national accounts comparison.
- United Nations (2008). *The System of National Accounts 2008*.
- Zuñiga-Cordero, A. (2018). Income inequality in costa rica: An inquiry towards distributional national accounts (dina). Master’s dissertation, Paris School of Economics, Paris.
- Zwijnenburg, J. (2019, 5). Unequal Distributions: EG DNA versus DINA Approach. *AEA Papers and Proceedings* 109, 296–301.
- Zwijnenburg, J., S. Bournot, F. Giovannelly, and F. Giovannelli (2017). Expert Group on Disparities in a National Accounts Framework: Results from the 2015 Exercise.