INSTITUTO FIGUEROLA
DE HISTORIA Y CIENCIAS SOCIALES

# Revealing the diversity and complexity behind long-term income inequality in Latin America: a new dataset, 1920-2011

Pablo Astorga

**Working Papers in Economic History**

**2021-11**
ISSN: 2341-2542

uc3m | Universidad Carlos III de Madrid

Serie disponible en http://hdl.handle.net/10016/19600

Web: http://portal.uc3m.es/portal/page/portal/instituto_figuerola

Correo electrónico: figuerola@uc3m.es

# REVEALING THE DIVERSITY AND COMPLEXITY BEHIND LONG-TERM INCOME INEQUALITY IN LATIN AMERICA: A NEW DATASET, 1920-2011

*Pablo Astorga*
Institut Barcelona Estudis Internacionals (IBEI)

ABSTRACT

The period between 1920 and 1980 is of great importance for the study of inequality in Latin America because of the occurrence of state-led, protected industrialisation amid structural, demographic and institutional transformations. Although there are valuable contributions at the country level, the study of income inequality from a broad regional perspective has been hindered by limitations of comparable metrics. To address this gap a new dataset has been assembled including Argentina, Brazil, Chile, Colombia, Mexico, and Venezuela. The approach adopted distinguishes four occupational groups: the top group includes employers, managers and professionals; the remaining three groups are defined according to the workers' skill level, largely receiving wage income. This allows for the calculation of inequality between and within groups, as well as overall Ginis for all income and wage income. The frequency of the series is annual, making it possible to track closely inequality trajectories. Despite being a high-inequality region, this new evidence reveals great diversity of outcomes across the six countries and complexity within the occupational structure. There is no single inequality metric that captures the whole story. Looking forward, this dataset opens the door to undertake econometric analysis to unpick the inequality contribution of key drivers such as the terms of trade and structural change.[1]

Keywords: income inequality, economic development, Latin America
*Clasificación JEL:* O54, O15, J31

Correspondence to: pastorga@ibei.org

---

[1] *This work is dedicated to Rosemary Thorp and Valpy FitzGerald whose outstanding contributions to the study of economic and social development in Latin America have been a source of learning and inspiration*.
I am grateful to Diego Castañeda, Eduardo Martín Cuesta, Ewout Frankema, Leonardo Maldonado, Marc Morgan, Eustáquio Reis, Javier Rodríguez Weber, Pedro Souza and Hector Valecillos for generously sharing their data and estimates and for guidance. Thanks to Nadine Kazerounian, Marc Morgan, Esteban Nicolini, Pedro Souza, Matthias vom Hau, Martijn Vlaskamp, Borge Wietzke and Henry Willebald for helpful comments. Special recognition to Branco Milanovic for illuminating observations. Erick Polin Castañeda provided effective research assistance during the last stage of this work. This research has received financial support as part of the project PGC2018-095821-B-I00: (MCIU/AEI/FEDER, UE).

## 1. INTRODUCTION

Imagine an inquisitive person who, looking through a blurry window, tries to form an idea of the landscape that once was. The onlooker can clear some areas to extend his view on the past. And the more he clears the glass, the sharper the picture that emerges. Thus, the researcher, sifting through historical data, can gather information to improve the quality and accuracy of the available evidence, separating historical fact from statistical illusion...or delusion. This work offers more pieces to advance the completion of Latin America's "inequality jigsaw puzzle". And in doing so, it also poses new questions for our understanding of the drivers of inequality in the region.

The paper introduces a new set of income inequality Ginis for Argentina, Brazil, Chile, Colombia, Mexico and Venezuela (LA6) during the period 1920-2011. These countries accounted for about three-quarters of the population of Latin America over the last century and thus are representative of the region as a whole. From a long-term perspective the decades between 1920 and 1980 are of particular interest and importance for the study of income inequality (and other inequalities) in the region because of the occurrence of state-led, protected industrialisation amid structural, demographic and institutional transformations. All of these changes are expected to have significant distributional implications, combining forces expected to be disequalising (i.e., the Kuznets-Lewis process) or equalising (the introduction of pro-labour institutions). Also, significant advances in schooling have the potential to curb the rise in inequality. However, although there are valuable contributions at the country level, the study of income inequality during this period from a regional perspective has been hindered by limitations of comparable metrics and a reduced sample of countries. This is particularly evident when compared with the more prolific literature and richer evidence available for the post 1980 years with frequent official household budget surveys largely comparable across countries.

The general picture of Latin America around 1920 was one of largely rural societies, poorly educated, with limited or incipient development of manufacturing, with economies dependent on the export of a handful of raw materials, and low rates of population growth. In response to the external shocks brought about by the Great Depression and the two World Wars, many countries in the region underwent major economic adjustments and revised their growth strategies. By the 1930s many economies turned more protectionist to promote domestic manufacturing. This move gave way to an explicit strategy of import-substituting industrialisation led by the state that dominated economic policy until the 1970s. This was accompanied by a process of rapid urbanisation (particularly in Brazil, Colombia, Mexico, and Venezuela), institutional transformations in the labour market and fiscal policy, mass education, high population growth, and economic diversification (Thorp, 1998). This period of industrialisation was followed by the Debt Crisis and the implementation of neoliberal reforms in the 1980s and the 1990s, and by the China-led commodity boom of the 2000s. These final decades, although with contrasting results in terms of economic growth and welfare, shared a process of de-industrialisation amid more open economies and rising labour informality in the service sector, with clear inequality implications.

What is the evidence available since 1920 to study the impact of these transformations?[2] Broadly speaking, the ensuing empirical study on income inequality can be divided into four strands. The first, social tables, is the more comprehensive approach to measure income inequality in periods where personal income surveys and fiscal data are limited. This method combines detailed data of benchmark years from population censuses with data on income from other sources (Milanovic et al., 2010; Allen, 2019). There are also dynamic social tables, where annual income data are used to fill the gap between benchmark years. In Latin America, there have been important efforts in constructing social tables in the last two decades or so. For instance, Castañeda and Bengtsson (2020) on Mexico, and on the dynamic variety, (Gómez León, 2021) on Brazil, (Rodríguez Weber, 2014) on Chile, (Rodríguez Weber, 2017 – based on Londoño, 1995) on Colombia, and (Bértola, 2005) on Uruguay. These detailed works based on official records, offer valuable insights, primarily on inequality levels at benchmark years and, depending on the case, on trends. Also, they pay special attention to the inclusion of property income as long as the data allows. However, there are also some limitations such as methodology inconsistencies across time periods, or data used in different benchmark years. Also, they are of limited comparability across countries, either because of differences in methodology and/or temporal span, and, therefore, are not well suited to offer a regional perspective on inequality.

[2] This is only a partial review of the empirical inequality literature on the region. I am only including those contributions more relevant to my work. For a more comprehensive review see Bértola and Willianson (2017) and references therein.

A second strand puts more emphasis on a multicountry scope the main evidence from which comes from wages and labour income. For instance, estimations of trajectories of labour income shares are offered by Frankema (2010) in Argentina, Brazil, and Mexico during the 20th century, and by Astorga (2017a) on those three countries plus Chile, Colombia, and Venezuela from 1900 to 2011. In another contribution, Frankema (2012) examines long-run industrial wage inequality in Argentina, Brazil, and Chile and concludes that aggregate inequality indicators (e.g., income Ginis) do not reveal much about the changing determinants of inequality, when the latter affect such indicators in opposite directions. And, therefore, it is necessary to include partial inequality indicators (i.e., inter-industry wage inequality, skill premiums) to help isolate the contributions of changing economic circumstances or political-institutional reforms. This finding is of particular relevance to my work.

The third empirical strand relies on household budget surveys (HBS) as the source of data to calculate personal or household inequality since the 1980s. There are earlier estimates for a handful of countries but they are not fully comparable across countries or across time (see Oscar Altimir estimates in Thorp, 1998, Statistical Appendix, some of which are included in Figure 1 below). Also, there is the recent work by Gazeley et al. (2018) on historical household surveys in Latin America 1913-1970. However, one well known limitation of the use of HBS in the region is the underestimation of top incomes, particularly property income. For instance, in a detailed study on household surveys in Latin America during the final two decades of the 20th century, Szekely and Hilgert (1999) found that top incomes were grossly underestimated - both because of the underrepresentation of rich individuals in the surveys and the underreporting of non-labour income.

This limitation leads us to the fourth strand based on the use of tax records. For recent decades they have been used to correct the underestimation of top incomes in the official household surveys and to produce more comprehensive income Ginis (Souza and Medeiros, 2015, Morgan and Souza, 2019 in Brazil; Burdín et al., 2019 in Uruguay;). Tax records also allow to track top incomes during periods without official HBS (Alvaredo, 2010 in Argentina; Flores et al., 2019 in Chile; Alvaredo & Londoño Vélez, 2013 in Colombia). However, problems of tax evasion and avoidance and fiscal data limitations together with methodological breaks limit the use of this approach to shed light on income concentration and inequality in Latin America over the long term.[3]

To widen the options available to study income inequality in the region, I adopt an estimating approach that largely relies on wage data, but that also makes allowances for non-labour income. To that end, I assembled a new set of annual Ginis based on four occupational skills categories, which can also capture developments on, grossly defined, functional inequality. These can be thought of as dynamic social tables with a reduced number of groups. The main data sources are population censuses (particularly, for the economically active population – EAP), wages reflecting various skill levels, and income aggregates from national accounts and the work of economic historians. The chosen methodology guaranties the comparability of inequality outcomes over time and across countries informing about commonality and diversity. In addition, the new series can unveil trajectories in income inequality at different levels of the structure of occupational groups, as well as of the between-group and within-group inequality components.

This new evidence reveals great diversity of outcomes across the six countries and complexity within the occupational structure. Confirming and extending Frankema's finding, there is no single inequality metric that captures the whole story. Within-group inequality largely shows contrasting patterns when compared to the between-group component; whereas, comparisons across countries are both mixed in direction and strength. Despite the common historical roots and structural and institutional features, diversity and complexity are two words that can well summarise the evidence. Overall, this evidence points to the combined action of drivers with different timing, opposite directions and varied intensity.

The remainder of the paper is structured as follows. It starts with the necessary explanation of the methodology and the estimation with particular attention to the new estimation of within inequality: Section 2 introduces the occupational groups and the calculation of various Ginis; Section 3 is devoted to explaining the procedure followed to estimate within-group inequality and the issue of income overlaps. Then Section 4 presents the series, compares them with alternative income Ginis, and highlights key patterns in the inequality trajectories across

[3] Jiménez et al. (2010) estimate average income tax evasion c.2005 equivalent to 4.6% of GDP in a sample of seven Latin American countries including Argentina, Chile, and Mexico. See also Alvaredo (2010) for concerns on the use of historical tax data in Argentina.

metrics and countries.[4] Section 5 concludes. Three annexes include complementary material. Annex A includes tables summarising labour shares and income ratios by lustrums, normality tests, calculations of income overlaps between the groups in benchmark years, and figures with country charts with all the six Ginis; Annex B offers details on the estimation procedure and assumptions made when calculating income aggregates and EAP shares; Annex C presents a detailed account of the procedure and the data sources used to assemble supporting series of wage dispersion for blue- and white-collar workers in manufacturing, as well as unskilled workers.[5]

## 2. Occupational Income Ginis

The starting point is the construction of dynamic distribution tables for the LA6 based on income estimates of four occupational groups, following the methodology in FitzGerald (2008). For each country, the economically active population (EAP) is divided into four groups: Group 1 (employers, managers, and professionals), Group 2 (technicians and administrators - white collar workers), Group 3 (semi-skilled blue collars workers, other urban workers in relatively low productivity sectors such as retailing and transport, and artisans), and Group 4 (rural workers and personal services – including domestic servants – plus unskilled urban workers). These groups are themselves an aggregation of the categories used in ECLAC's annual publication Panorama Social. To ensure consistency with the overall EAP series, the labour force in Group 3 is calculated as a residual.[6] The groups' sizes change over time in response to developments in skills formation, demography, and living standards (Astorga et al., 2005). The distribution of income per occupational category in a given year is defined as:

$$(1)\ \sum_{i=1}^{4} e_i r_i = 1,$$

where $e_i$ is the EAP share of group $i$, and $r_i$ is the ratio of the mean income of group $i$ to the mean income for the EAP as a whole (i.e., income per person engaged). The income share of each group ($s_i$) is equal to $e_i r_i$.

[4] It is beyond the scope of this work to offer discussion of particular inequality-related developments in the six countries. Astorga (2017b) and Arroyo & Astorga (2017) offer commentary on trajectories in between-group inequalities. But country stories are largely a pending task, particularly regarding within-group inequality.

[5] The dataset will be released with the journal version of the paper.

[6] See Annex B2 for more details on the calculation of the four labour shares.

The overall measure of income per person engaged reflects, where possible, the pre-fisc personal income concept of the national accounts. I am choosing this concept rather than net national income to avoid an overestimation of the income share of Group 1 that would result if items such as the net surplus of the public sector, and indirect and corporate taxes were included. Although, since the 1980s, there is enough data to account for net taxes, this is more problematic for the previous years. In any case, there was limited redistribution via direct transfers in the region during most of the 20th century (Goñi et al., 2011) and the analysis of the series pre-fisc or post-fisc should lead to similar conclusions. Also bear in mind that I am not considering the distributive impact of social spending (e.g., health and education) which has risen throughout the region since the 1980s, though exhibiting high volatility and following the swings in economic activity.[7]

Ideally, income estimates should make allowances for the subsistence economy. However, there is little systematic and consistent evidence of its size (particularly important in the early decades of the 20th century), which could be used to make an adjustment (Berg, 1970). To the extent that the population in the subsistence sector is included in the census, I am assigning them an income equal to the unskilled wage. Regarding mixed income, I assume that the earnings of the self-employed in the lower three occupational categories are largely made up of labour income, and that they can be approximated by the corresponding average wage in each category (Amarante et al., 2014). Also, difficult to obtain for most of the period are differences in employment levels across occupational groups. My calculation in each category assumes full-time pay rates and that unemployment was affecting all categories equally.

The income share for Group 1 ($s_1$) is calculated as a residual by subtracting the income shares for the other three groups:

(2) $s_1 = e_1 r_1 = \{1 - \sum_{i=2}^{4} e_i r_i\}$.

This share is likely to capture most of the property income (distributed profits, dividends, rents and interest payments) for all the EAP, together with labour income of managers and

[7] On balance, whereas the evidence of the 2000s shows the implementation of progressive social spending, the available historical estimates for a handful of countries during the last century points to a diverse distributional impact across countries and periods, with relatively more progressive social interventions in Argentina. See Arroyo Abad and Lindert (2017).

professionals.[8] Natural resource rents - particularly important in Chile and Venezuela during most of the period - are included to the extent that they are reflected in household or personal income, but not when they were used to finance publicly provided services.

Because of the way it is calculated, $s_1$ may be potentially subject to a significant margin of error. However, in general, my estimates for the mean income of this group in the first half of the 20th century are consistent with data available on top earners. Also, when the data allows for a comparison, trends in $s_1$ are broadly consistent with the income share of gross profits in the national accounts (Astorga, 2017b, Table 2 and Figure A2). To estimate mean income of the remaining three occupational groups I rely on wage series assembled to reflect differences in skills (Astorga, 2017a, online annex).

Income shares are estimated using a combination of sources. Overall income figures come from the national accounts and, in the early decades, from the work of economic historians; the distribution of the labour force is sourced from population census or employment surveys; and wage data is largely collected from official statistics (using primary and secondary sources). In all of these sources the data is not self-reported and, thus, free from the well-known systematic under-reporting of property and self-employment income in modern household surveys. Finally, in the adopted procedure there is a need to conciliate the overall income data from the national accounts with the wage data compiled for the three lower occupational categories. This is described in Annex B1.

I use the Gini coefficient to measure income inequality for the four occupational groups, calculated as follows:

$$(3)\ G4 = G4B + G4W,$$

where G4B stands for between-group inequality and G4W for within-group inequality.

G4B is the inequality that would be obtained if everybody in a given group was given the mean income for that group. It is calculated with the groups' mean incomes and their corresponding EAP shares, with groups ranked by their mean incomes in ascending order in a given year:

[8] The long-term evidence in developed economies (Piketty, 2014) shows that property income tends to be concentrated in the top 10% income group. And, in all probability, this is also true in Latin America owing to a historically high concentration of assets (Frankema, 2009).

(4) G4B = $\sum_{i=2}^{4}\sum_{j=1}^{i-1} e_i e_j |r_i - r_j|$,

where, as previously, $e_i$ is the EAP share of group *i* and $r_i$ is the ratio of the mean income of group *i* to that for the EAP as a whole.

G4W is a weighted sum of the Gini coefficient each group would have if it were a separate population ($G_i$). The contribution to within-group inequality of Group 1 and those of the three lower groups are presented separately because their estimation procedures are different (see below):

(5) G4W = $e_1 s_1 G_1 + \sum_{i=2}^{4} e_i\, s_i G_i$.

Because lack of micro data I do not adopt the traditional decomposition approach (e.g., Lambert and Aronson, 1993) that includes a term for residual inequality reflecting any income overlaps between groups. Therefore, G4 is a gross occupational Gini rather than the overall Gini that would result if the population were perfectly sorted by income (Modalsli, 2015). The potential implications of income overlap are discussed in Section 3.4.

I also construct narrower Ginis for the three lower occupational categories based on similar equations as in (3) to (5). These Ginis are gross overall (G3), between-group (G3B), and within-group (G3W). Having Ginis based on the four categories as well as on the three lower ones is of interest because the dominant forces affecting labour and property income are different. The former is driven by demand and supply conditions in the labour market (and, in turn, influenced by technology and skills formation), as well as by labour-market institutions (e.g., unions) and regulations on wages (e.g., minimum wage), whereas property income is primarily driven by factors such as savings and investment decisions, inheritance laws, and the rate of return to wealth.[9] And comparing them makes it possible to assess the extent to which the forces shaping property and labour income are acting in a reinforcing or offsetting manner at different points in time.[10] Group 1's high-income earners are likely belong to the economic elites, and its income share can shed light on their influence on inequality developments.

---

[9] Indeed, these differences in the main drivers are reflected in a relatively low average correlation between G4 and G3 equal to 0.25 in the LA6 between 1920 and 2011 (see Table 1 below). Paired correlations are calculated using a five-year period data, so as to minimise any distortions caused by interpolation. Also, outside the region, Gómez-León and Gabbuti (2021) estimated overall Ginis and labour Ginis using the same methodology to calculate the income share of proprietors in Italy between 1900 and 1950, and also found significant differences in trajectories.

[10] The inequality impact via property income under structural change is complex. In general, the combination of protected industrialization with a decline in agriculture generated winners (the new industrialists) and losers (the traditional landlords). Also, the state became a key economic actor via state-owned enterprises. See Rodríguez

## 3. Within-group inequality

A comment on the history of this research is in order. In the first stage, I assembled series for mean wages for the three lower categories and estimated the mean income of Group 1 as a residual (see eq. 2). With this information, I calculated series G4B and G3B (the latter driven by skill wage premiums), which were the attention of previous publications (Astorga, 2015, 2017b; Arroyo and Astorga, 2017).[11] At that time, I was unable to find enough data to inform consistently, and with a satisfactory coverage, on within-group inequality for the whole period. Some years later I had another look at this estimation challenge.[12] By then there was a much-improved online availability of official publications with historical wage statistics (e.g., on industrial surveys), together with new contributions to the inequality literature that eased my task. Data demands were also reduced significantly by starting the estimation in 1920 rather than in 1900 as in the first stage. This made it possible to assemble proxy series for wage dispersion that offered a reasonable match (particularly on changes) to the "true" (and unknown) income dispersion of my three lower occupational groups. The rest of this section describes this second stage of the research.

### 3.1. Wage dispersion

I assembled new series on wage dispersion measured by the coefficient of variation (*cv*) with a sufficient number of benchmark observations over the 1920-2011 period to capture underlying trends in changes in within-group income dispersion for the lower three occupational groups. Income dispersion for Group 1 is discussed in Section 3.3. To improve consistency and coverage, priority was given to data sources available to all six countries during most of the period covered. Here I present a summary of this task. See Annex C for full details by country.

For Group 4 (unskilled workers), I calculated wage dispersion across low-skilled occupations using wage data from various official statistical publications at a country level, ILO's October Inquiry including data on all six countries – but with uneven coverage, and from social tables compiled by economic historians (e.g., Chile and Mexico). For the later decades,

---

Weber (2015) for the analysis of Chile. Also, Allen (2019) for a discussion on these offsetting forces in property income during the industrial revolution in England.

[11] The present work also provides full estimations and source details for the previous publications. However, here I use household income as the overall income measure rather than net national income. However, the change in the income concept does not affect the conclusions drawn from the evidence.

[12] A silver lining of the Covid-19 lockdown…

when needed, I use income dispersion calculated from the centile structure of HBS at the lower end of the distribution of non-zero incomes (e.g., from the centile 1 to centile 35). Accounting for the rural-urban divide is a key estimation issue for this group, as I am covering a period where the region underwent a rapid process of internal migration. This is of particularly relevance in Brazil, Colombia, Mexico and Venezuela where the urbanisation rate went from under 20% in c.1920 to about 70% in c.1980 (Astorga et al. 2005). When data is available, I am assembling a representative sample of unskilled wages in both rural and urban activities in benchmark years with the proportions defined in line with the corresponding urbanisation rate at the time (for an example see Brazil and Mexico in Annex C).

To gauge income dispersion in Group 3 (dominated by semi-skilled workers) and Group 2 (relatively skilled workers), I am largely relying on official industrial censuses and surveys for blue-collar and white-collar workers in manufacturing according to the International Standard Industrial Classification (ISIC) with a breakdown by divisions (two-digits).[13] Ideally, the coverage should include other sectors such as construction, commerce and the public sector. However, manufacturing is the only sector with sufficient data across all six countries over the period of analysis. In addition, the wage data in manufacturing separate blue- and white-collar workers, which is crucial for constructing comparable and consistently defined proxy series for the two occupational groups. Therefore, I am assuming that changes in wage dispersion of blue and white-collar workers in manufacturing offer a reasonable proxy for wage dispersion in my middle groups, particularly in a period dominated by industrialisation.[14] The comparative evidence presented in Section 4 indicates that this conjecture results in Ginis the trajectories of which are broadly consistent with alternative Ginis.

However, the matching of the corresponding skill level is an issue that needs attention. The blue-collar category includes a proportion of unskilled workers, especially in industries such as food and textiles, that does not belong to Group 4; and some relatively skilled workers that would be better placed in Group 2. Meanwhile, the white-collar category includes salaries of managers and professionals, which belong to my Group 1, as well as some relatively low skilled (and paid) clerks that would be better placed in Group 3. Thus, in both cases the direct

[13] The data up to the 1980s usually refer to ISIC1 (including up to 20 industries), and to ISIC2 (up to 28 industries) thereafter.

[14] In his study of the Brazilian labour market Bacha (1979) uses blue-collar wages as representative of semi-skilled urban wages.

use of wage dispersion in blue- and white-collar categories would lead to an overestimation of the level of wage dispersion in Groups 3 and 2.

To address this problem, a downward adjustment to the dispersion level of both industrial categories is needed prior to using them as proxies for income dispersion in the two middle groups. Fortunately, there are some data with a similar industry breakdown that offer an indication of the magnitude of such an adjustment. Shipley (1977) has blue-collar workers in ten manufacturing industries in Argentina during the 1920s, separating unskilled and semi-skilled workers. This makes it possible to calculate wage dispersion for blue-collar workers with or without the unskilled. On average, the dispersion without the unskilled is about 0.87 of that of the whole blue-collar category. In a more recent period, a similar calculation for the period 1986-1991 with Argentina's wage data in manufacturing (ISIC2) gives a ratio close to 0.80 (ILO YLS 1996). As for the adjustment to the white-collar category, industrial censuses in Mexico in 1935, 1940, and 1945 (DGE, 1953) present income data separating directors and managers from other white-collar employees. On average, the dispersion in salaries for white collars without the directors and managers is about 0.80 of that of the whole white-collar category. Based on these calculations, I downscale blue- and white-collar wage dispersion by 0.85 ($adj_{bc}$) and 0.80 ($adj_{wc}$) respectively over the whole period. Wage dispersion for unskilled workers ($cv_{unsk}$) is left unadjusted, as in this case there is no skill mismatch.[15]

### 3.2. Ginis for the three lower occupational groups

The dispersion for the lower three wage-based groups (*cv'*s) are derived from the adjusted coefficients of variation of the series of white' and blue-collar wages and unskilled wages as follows: $cv'_2=adj_{wc}cv_{bc}$ ; $cv'_3=adj_{bc}cv_{bc}$ ; $cv'_4=cv_{unsk}$. These *cv's* are then used to estimate standard deviations compatible with the mean incomes of Groups 2, 3 and 4 obtained in the first stage. In each group, and in a given year, this is calculated as:

(6) $\sigma_i = cv'_i * u_i$; $i$=2 to 4, and where $u_i$ is the group *i* mean wage form the first stage.

---

[15] These adjustments affect the dispersion level (coefficients of variation) of the two manufacturing workers' categories across the board. Their main purpose is to provide better proxies for wage dispersion in my occupational groups, and to correct an otherwise excessive income overlap between them (see Section 3.4.). The downscaling does not affect the trajectories of my Ginis, nor the conclusions drawn from them.

This information can be used to simulate a Pen's income parade (Pen, 1971) in each group and year, assuming a given income distribution function (Modalsli, 2015). It is well-known that the whole income distribution is well fitted by a log-normal distribution with a Pareto upper tail. However, it is a moot point whether this is also true for different groups within a given population (e.g., unskilled workers or blue-collar workers).[16] To clarify this empirical issue, I performed normality tests[17] on a representative sample of the wage data available for my three lower occupational groups from industrial and occupational surveys in benchmark years. Results are summarised in Table A3. In most cases the null hypothesis of normality cannot be rejected. The evidence for the more limited unskilled wage data is also dominated by normality, though here there are more rejections of the null hypothesis.

Equally, when performing the same tests to a selection of perfectly-sorted quantiles (or income groups) in the HBS centile distributions - excluding zero incomes, normality tends to reflect well the income distribution that exclude the top ten centiles. For example, for centiles 1 to 35 (c1-35), c36-70, and c71-90.[18] However, normality, as expected, is rejected in most cases in c71-95 and in all cases for entire centile distributions (not shown). These results indicate that imposing a normal distributed income structure in each of my wage-based groups is a reasonable assumption when estimating the associated Lorenz curve.

Finally, to calculate Gini coefficients for each of the three lower occupational groups, each group's EAPs is divided into 25 quantiles (N=25), and the corresponding income ratios estimated with the use of a normal distribution.[19] The respective Ginis ($G_i$) for a given year are calculated as follows:

(7) $G_i = \sum_{j=2}^{N} \sum_{k=1}^{j-1} e_j e_k |r_j - r_k|; \quad i = 2 \text{ to } 4.$

[16] In his analysis of micro data on pre-industrial social groups Modalsli assumes log-normality as the preferred probability distribution function. This is based on the evidence provided by data on three cases of pre-industrial societies: Tuscany in 1427, using wealth data; Bihar in 1807, using expenditure data; and Norway in 1868 using income data for the upper 33% of the population. All three cases tend to include people at the top of the distribution, which suits the use of log-normal distribution.

[17] A total of four tests are calculated: Shapiro-Wilk, Anderson-Darling, Lilliefors, and Jarque-Bera using XLSTAT. Of these, the Shapiro-Wilk test tends to have the better power in relatively small samples (e.g., fewer than 100 observations). See Yap and Sim (2011).

[18] These income groups are chosen to reflect roughly average values of the EAPs centiles in my three lower groups in the 1990s in the LA6 (see Table A1).

[19] Calculations are done in Excel using the NORMINV function with three parameters: the accumulated EAP share with increments of four percentage points (=1/25*100), and the mean income and standard deviation of each group in a given year.

### 3.3. GINIS OF GROUP 1

For the top group it is not possible to adopt the same estimating procedure used in the other groups because of the lack of detailed within-group income data for most of the countries and period. Facing this limitation, my main purpose here is to estimate the within-Group 1 inequality level at some benchmark years, and to gauge a plausible trajectory during the rest of the period. In this way, I can complete the calculation of G4W (eq.5) and G4 (eq.3). My starting point is data on the distribution of income of the top 10 centiles from official household budget surveys in years when the EAP of Group 1 was close to 10%. The benchmarks are 1981 in Brazil, 1992 in Chile, Colombia in 2007, and 1989 in Mexico. For Argentina and Venezuela, I assume an income Gini of 0.50 in those years when their Group1's EAP was at 10% (1989-90 and 1976-77 respectively). The assumed value is the simple average of the estimated Ginis for the other four countries in years with EAPs close to 10%. There are data on the top centiles in Venezuela, but these only include labour income (Maldonado, 2021).

Because of the underestimation of income of high earners in the HBS, it is necessary to make an adjustment to better reflect the level of within-group inequality. This is done by boosting the income for the 100$^{th}$ centile, so that the ratio of the top1% to the top10% income shares equals 0.45. Such an adjustment factor is in line with an average of a similar ratio calculated from Pedro Souza´s fiscal data for Brazil in benchmark years between 1981 and 2006.[20] With the top ten EAP centiles and the adjusted income ratios, the $G_1$s are calculated as:

$$(8)\ G_1 = \sum_{j=2}^{N} \sum_{k=1}^{j-1} e_j e_k |r_j - r_k|; \quad \textit{with } N = 10 \textit{ and } e_j = e_k = 0.1.$$

To estimate changes in the top-group inequality backwards to 1920 I use auxiliary country Gini series calculated as in (7) but with changing EAP shares (see Table A1) with N $\leq$ 10, and using the corresponding average income ratios calculated from the HBS centile data available in the 1980s and 1990s.[21] Basically, I am assuming that inequality changes within the top

[20] Piketty (2014, 292) estimates the same ratio for the US in the 2000s at around 0.4. The HBS top1% to top10% average share during the same period is around 0.30 in Brazil (IBGE PNAD), as well as in Chile (1992-2006) and Mexico (1992-2008).

[21] In Argentina, Chile, Colombia and Venezuela I use Chile's HBS in 1992, 1996 and 1998 sourced from LIS; in Brazil the country's HBS in 1981, 1985, 1989 from IBGE's PNAD; in Mexico HBS in 1984, 1989, and 1992 from INEGI. In Colombia during the period 1948-1986 the auxiliary Ginis grow in line with income Ginis for "landlords"

group prior to 1980 or so were similar to those calculated with HBS data with matching centile structure in a more recent period.

### 3.4. Income overlaps across groups

Before presenting the inequality series, I add some discussion on the issue of income overlap and the extent to which the four occupational groups offer an appropriate breakdown of the EAP that minimizes the potential for overlapping. A useful concept is that of 'well-apportioned' groups. For a group to have a separate identity the income differences within the group should be less than the differences across the groups, and the weighted sum of within-group Ginis should not be larger than the between-group Gini (Modalsli, 2015; Milanovic et al., 2010). Under this concept, my four occupational groups are well-apportioned, and this should translate into limited income overlaps between groups.

Table A4 includes a summary of income estimates at different points of the distribution of Groups 4,3, and 2 in five benchmark years (from 1920 to 2000) to inform about the extent of income overlaps. They are calculated based on the so-called three-sigma rule of thumb (68-95-99.7 rule). This rule states that for a normally distributed variable 68%, of all values lie within one standard deviations of the mean ($u \pm \sigma$), 95% within two standard deviations ($u \pm 2\sigma$), and 99.7% within three standard deviations of the mean ($u \pm 3\sigma$). The overlap between Groups 4 and 3 is limited and largely affecting the EAP above $+1\sigma_4$ and below $-1\sigma_3$, involving the 16% (13.5% + 2.5%) top end and lower end of the two group's income distributions. The overlap is most significant for values above $+1\sigma_3$ and below $-1\sigma_2$. By contrast, because of large mean-income ratios (see end column "$u_1/u_2$") in all six countries, the overlap is likely to be minimal between Group 2 and Group 1.[22] Because its distribution is not normal, it is not possible to apply the three-sigma rule to the top group.

---

and "capitalists", weighted by their respective income shares (Rodríguez Weber, 2017). Any gaps are filled with linear interpolation.

[22] There are two pieces of additional evidence to support this claim. Detailed social tables in Mexico in 1930 and 1940 show income ratios between equivalent Group 1 and Group 2 of 13 and 6.3 respectively (Castañeda and Bengtsson, 2020). And according ECLAC's Panorama Social (2000), based on data around 1997 for eight Latin American countries (including the LA-6 but Argentina), Group 1's mean income was about 3 times higher than that of Group 2, and the mean income of the lower sub-group in Group 1 (managers) was 2.2 times higher than that of the higher sub-group in Group 2 (technicians). Note that, because of the top-income underestimation in the household surveys, these ratios should be taken as lower-bound values.

In order to gauge the impact of income overlaps on the occupational Ginis, and in particular on G3W - the metric where the effect is most significant, I reduced the standard deviation of Groups 3 and 2 by 25% uniformly over the period in the six countries. Such an adjustment minimises income overlaps across the three lower groups and produces near perfectly sorted groups (results not shown).[23] As expected, lower dispersions in both groups reduces average inequality levels over the period, especially of G3W. For instance, in the case of G4, G3 and G3W by 0.9%, 4.8% and 23.1% respectively in Argentina, by 0.7%, 2.8% and 11.6% in Brazil, and by 1.2%, 5.6% and 21.5% in Chile. However, Gini trajectories are largely unaffected.

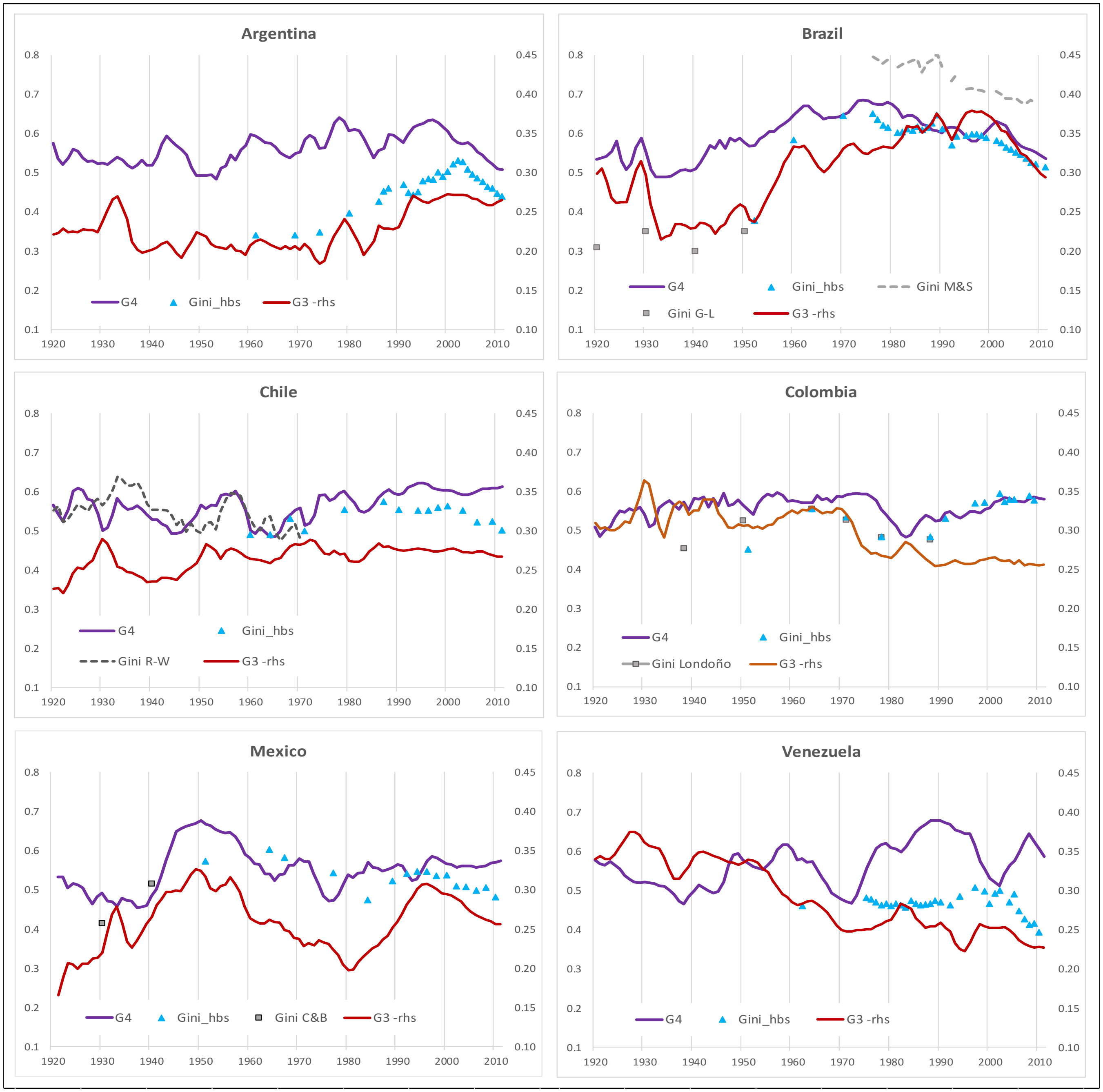
## 4. The Inequality Evidence

This section starts by comparing my overall Ginis (G4 and G3) with alternative income Ginis available for the period under analysis. Such comparisons offer information to check consistency across available inequality estimates (particularly on trajectories), as well as to assess the feasibility of the new series. This is followed by discussion on paired correlations across metrics and countries. The comparative evidence is presented in Figure 1, including social-tables Ginis estimated by economic historians and HBS Ginis for the more recent decades.[24] Figure A1 (all four groups) and Figure A2 (the three lower groups) in Annex A include charts with all the new occupational Ginis series by country.

A visual inspection of the charts indicates broad consistency in inequality patterns. In Brazil it is possible to make a comparison with Gómez León (2021) social-tables benchmarks in 1930, 1935, 1940 and 1950. Although the levels of her Ginis are lower, the trends are broadly matching those in G4, with a rise between 1920 and 1930, then a fall to 1940 and a rise to 1950. For this country, a second comparison can be made with the Gini series of Morgan and Souza (2019), which combines tax and HBS data from 1976 to 2010.[25] Their Gini and my G4 display a matching downward secular trend. In Chile, the trajectory of my G4 series is in line

[23] The resulting overall Gini is close to the concept of Gini2 in Milanovic et al. (2010), where in addition to between-class inequality some within-class inequality is included under the strong assumption that all members of a given social class are poorer or richer than those respectively above or below them.
[24] Since 1990 there are also available Ginis that only include labour income (not shown). Their trajectories are closely in tune with the all-incomes HBS Ginis. See ECLAC website.
[25] I use their series for "individualistic adults" which includes all adults aged 20 and over, taking the income reported by each individual separately. Of their various Ginis, this is the one that uses an income concept closer to personal income.

FIGURE 1: OVERALL OCCUPATIONAL GINIS AND ALTERNATIVE INCOME GINIS




All G4 and G3 (plotted on the right-hand side) series are three-years moving averages. In Brazil, Gini G-L from Gómez León (2021), and Gini M&S from Morgan & Souza (2019) which combines tax and HBS data (all individual adults aged 20 and over); in Chile, Gini R-W from Rodriguez Weber (2014); in Colombia, Gini Londoño from Londoño (1995); in Mexico, Gini C&B in1930 and 1940 from Castañeda & Bengtsson (2020). Sources for Ginis_hbs: Argentina 1976-2011 CEDLAS (urban data), to go back to 1960 it uses Altimir compilation in Thorp (1998, Statistical Appendix); Brazil 1976-2011 IBGE PNAD and Altimir to go back to early 1950s; Chile 1990-2011 ECLAC and Altimir for Greater Santiago to go back to 1960; Colombia 1988-2009 Szekely & Sámano (2012) and Altimir to go back to the early 1950s; Mexico 1994-2010 ECLAC, 1984-1994 from Szekely & Sámano (2012) and Altimir to go back to early 1950; Venezuela 1990-2010 ECLAC, and Baptista (1997) to go back to 1962.

with that of the Gini series of Rodriguez Weber (2014) between 1920 and 1971. This is of no surprise because the latter work is my main source for wage data for Chile. However, importantly, this also shows that the use of only four “well-apportioned” occupational categories can do a good job at capturing the evolution of income inequality estimated with a much greater level of disaggregation. In Colombia, Londoño (1995) offers income Ginis in six benchmark years between 1938 and 1988, using a combination of national accounts and employment and household surveys. Here, although his Gini and my G4 have rising trends between c.1940 and c.1970, the G3 trajectory offers a better overall match. Finally, in Mexico the income Ginis of Castañeda and Bengtsson (2020) in 1930 and 1940 show rising inequality between the two benchmarks; a trend that is also present in both my G4 and G3 series.

Comparison of trajectories can also be made with the official, all income, HBS Ginis (Gini_hbs in the charts). Whereas trends in my G4s tend to match those of social-tables Ginis, there is a weak synchronicity with the Gini_hbs. A result that points to a differentiating effect of property income in shaping the overall inequality outcome.[26] By contrast, trajectories in G3s (plotted on the right-hand side scale) and Gini_hbs are broadly in tune in all six countries, particularly in Argentina, Brazil, Chile and Mexico. As would be expected, G3s’ levels are much lower than those calculated from household surveys (for instance, because of the exclusion from the former of labour income for professional and managers). Although my G3s and the HBS Ginis are constructed differently (e.g., changing vs. fixed EAP shares; not perfectly sorted data vs. perfectly ordered) and use different data (records at work versus surveys at home), in both cases labour is the main source of income, which should provide a common ground for co-movements in both metrics. Also, notice that, although my pre-fisc series exclude the impact of redistribution policies in the 2000s (e.g., conditional cash transfer programmes), the main driver of the inequality decline in the region during that decade was strong growth in labour income for low-skilled workers (Azevedo et al., 2013) – which is accounted for in my Ginis.

[26] In Brazil there is also weak synchronicity between Gini_hbs and the more encompassing Gini of Morgan and Souza (2019).

### 4.1. PAIR CORRELATIONS BY METRICS AND COUNTRIES

Having a variety of inequality measures looking at components and group structure for each of the six countries allows for a richer analysis of diversity and commonality. Here I present pair correlations between occupational Ginis using observations every five years, so as to minimise distortions caused by the use of interpolation in the underlying EAP series. First, in Table 1, between the various Ginis in a given country to assess the extent to which the inequality components acted in a reinforcing or offsetting manner. And, secondly, in Table 2, between a given occupational Gini across countries to inform about the extent of synchronicity or asynchronicity in trajectories.

TABLE 1: PAIR CORRELATIONS BETWEEN METRICS BY COUNTRY

| | **ARG** | **BRA** | **CHI** | **COL** | **MEX** | **VEN** | **LA6** |
|---|---|---|---|---|---|---|---|
| **G4 & G3** | 0.27 | 0.63 | 0.27 | 0.25 | 0.57 | -0.48 | *0.25* |
| **G4B & G3B** | -0.01 | 0.69 | 0.25 | 0.31 | 0.61 | -0.38 | *0.24* |
| **G4 & G4B** | 0.95 | 0.99 | 0.96 | 0.99 | 0.99 | 0.99 | *0.98* |
| **G3 & G3B** | 0.89 | 0.97 | 0.63 | 0.98 | 0.98 | 0.99 | *0.91* |
| **G4 & G4W** | 0.39 | -0.51 | 0.22 | -0.16 | -0.38 | 0.15 | *-0.05* |
| **G3 & G3W** | 0.66 | 0.02 | 0.64 | 0.24 | 0.18 | 0.83 | *0.43* |
| **G4B & G4W** | 0.09 | -0.62 | -0.08 | -0.32 | -0.49 | 0.05 | *-0.23* |
| **G3B & G3W** | 0.24 | -0.21 | -0.19 | 0.03 | 0.03 | 0.74 | *0.11* |

Correlations based on a five-year period data. LA-6 is the simple average of the six countries.

There are some patterns to highlight:

- When looking at comparisons between Ginis including all four occupational groups and those including the three lower groups - G4s & G3s and G4Bs & G3Bs - there is a relatively low correlation (higher in Brazil) and positive (except Venezuela), indicating the differentiating role of high-earners income in shaping inequality trajectories. Therefore, the analysis of the impact of potential inequality drivers needs to take this into account, as the explanations suitable to the lower three groups are likely to be insufficient to shed light on the more encompassing Ginis. This result also has implications for the use of HBS Ginis, whose trajectories are largely in tune with those in G3s (as shown in Figure 1), and can only offer a partial view by largely reflecting the action of drivers primarily affecting labour income.

- Regarding comparisons between the overall and the between components with the same group structure, changes in the overall occupational Ginis (G4 and G3) are strongly correlated with their corresponding between-group Ginis (G4B and G3B), indicating that trajectories in between-group inequality, and the factors that influence them, dominate the behaviour of the overall Ginis and that the inclusion of within-group inequality in itself, has a limited capacity to affect aggregate inequality.[27]
- Comparisons of pair correlations involving the three components in each of the four occupational groups (G4s & G4Bs paired with G4Ws) show, in general, low correlation values and with mixed signs. In particular, there are noticeable contrasting trajectories in Brazil and Mexico, indicating that there are different partial inequality stories to be told, a fact that would be ignored if attention were only placed on the overall inequalities (G4 and G3) dominated by developments in the between components (G4B and G3B). By contrast, similar comparisons with the same components but in the three wage-based groups show mostly positive correlations, particularly between G3 & G3W in Argentina, Chile and Venezuela. These reinforcing moves are reflecting dispersion dynamics in more homogeneously defined workers' occupations (excluding the top group) affected by labour regulations. All in all, this outcome is of particular interest for the study of the inequality impact of industrialization in the region. Although this deserves further investigation, this evidence suggests that the action of inequality drivers (market or policy-driven) had different effects at different inequality components and across the group structure.[28]

When pairing countries, the more encompassing overall G4 shows, in general, low and positive correlations; whereas although the narrower G3 shows stronger associations, they are dominated by asynchronicity. Both results, especially the second, point to country diversity in inequality. This is somehow unexpected, given the commonality across these countries in terms of historical roots, their insertion in the world economy, and similarities in structural and institutional transformations. The correlation patterns observed in the two overall Ginis are

[27] The dominance of between-group inequality in overall inequality is also reported in Milanovic et al. (2010) in their inequality analysis of social tables in pre-industrial societies.

[28] Notice that because income inequality in Groups 2 and 3 reflect wage dispersion in white- and blue-collar workers in manufacturing respectively, this result is particularly influenced by developments in manufacturing, hence, industrialization.

largely concurrent with those in the corresponding between-group Ginis. A result which is consistent with the fact that overall inequality is driven by the between-group component.

When looking at specific country pairs, the correlations between neighbouring Colombia and Venezuela stand out. They are relatively high and negative in G4 and G4B (-0.55 and -0.50 respectively), and high but positive (synchronised) in G3 and G3B (0.77 and 0.65). Therefore,

TABLE 2: PAIR CORRELATIONS ACROSS COUNTRIES BY METRICS

| ***Overall G4*** | | | | | | ***Overall G3*** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | BRA | CHI | COL | MEX | VEN | | BRA | CHI | COL | MEX | VEN |
| ARG | 0.48 | 0.17 | -0.37 | -0.08 | 0.35 | ARG | 0.44 | 0.30 | -0.57 | 0.17 | -0.45 |
| BRA | | 0.04 | 0.01 | 0.20 | 0.29 | BRA | | 0.68 | -0.62 | 0.04 | -0.80 |
| CHI | | | -0.17 | 0.05 | 0.44 | CHI | | | -0.31 | 0.38 | -0.58 |
| COL | | | | 0.03 | -0.55 | COL | | | | -0.19 | 0.77 |
| MEX | | | | | 0.16 | MEX | | | | | -0.22 |
| ***Between-groups G4B*** | | | | | | ***Between-groups G3B*** | | | | | |
| | BRA | CHI | COL | MEX | VEN | | BRA | CHI | COL | MEX | VEN |
| ARG | 0.42 | -0.06 | -0.29 | -0.13 | 0.21 | ARG | 0.03 | 0.31 | -0.31 | -0.12 | -0.06 |
| BRA | | -0.19 | 0.16 | 0.26 | 0.27 | BRA | | 0.17 | -0.38 | 0.09 | -0.74 |
| CHI | | | -0.24 | -0.08 | 0.27 | CHI | | | 0.13 | -0.13 | -0.02 |
| COL | | | | 0.15 | -0.50 | COL | | | | -0.08 | 0.65 |
| MEX | | | | | 0.15 | MEX | | | | | -0.26 |
| ***Within-groups G4W*** | | | | | | ***Within-groups G3W*** | | | | | |
| | BRA | CHI | COL | MEX | VEN | | BRA | CHI | COL | MEX | VEN |
| ARG | 0.58 | 0.78 | 0.11 | 0.62 | 0.59 | ARG | 0.46 | 0.58 | -0.20 | 0.27 | -0.53 |
| BRA | | 0.13 | 0.62 | 0.82 | 0.72 | BRA | | -0.20 | 0.57 | 0.81 | -0.08 |
| CHI | | | -0.25 | 0.24 | 0.25 | CHI | | | -0.76 | -0.28 | -0.62 |
| COL | | | | 0.66 | 0.13 | COL | | | | 0.62 | 0.51 |
| MEX | | | | | 0.50 | MEX | | | | | 0.05 |

Correlations based on a five-year period data.

although inequality trajectories in wage inequality are largely in tune in both countries (though the sustained shared decline started in the early-1950s in Venezuela and around 1970 in Colombia), the income accruing to the top occupational group has a contrasting and dominating behaviour in the broader G4 (see Figure A1). This result is likely to reflect, on the one hand, the differentiating factor of inequality dynamics of oil rents in Venezuela and, on the other, the apparent similarities in developments in the labour market and in the timing of protected industrialization. Synchronised trends also dominate in the within-group labour component, particularly before the mid-1980s (see Figure A2).

By contrast, correlations between Chile and Venezuela show moderately concurrent trajectories in G4, G4B and G4W, but strong and negative associations in G3, G3B, G3W. Thus, in this case, there is more synchronicity of moves in the Group1's income, but different inequality stories arising from the labour market. Other country pairs add to the diversity of outcomes. For instance, Argentina-Brazil and Argentina-Colombia, both G4 and G3 show coinciding signs, but synchronised in the former and desynchronised in the latter.

Of the two inequality components, the within-group Ginis show the stronger country correlations: mostly positive in G4W, and with mixed signs in G3W. The prevalence of synchronised movements in the former is partly the result of the assumptions used in the construction of income dispersion in Group 1 (see Section 3.3.). Therefore, I focus the discussion on within-group inequality in the lower three groups. There are eight correlation values above ±0.50 out of a total of 15; 5 are positive and 3 negative. Country patterns in G3W are not replicated in G3B, which points to the presence of different dynamics in the wage-based groups. On the one hand, in the premiums between occupational groups and, on the other, in wage dispersion within the groups. Here as well, the evidence suggest variation in the underlying forces such as demand and supply for skills, policies and institutions in the labour market, and the timing and nature of urbanisation and industrialization.

All in all, pair correlations across countries are both mixed in direction and strength (low in G4, higher in G3W). Diversity dominates, implying that there is limited support for a regional inequality pattern in this sample of countries. Indeed, the main pattern seems to be absence of one.

## 5. Conclusions

Building on previous work, this paper presents long-term series of income inequality in Latin American during the period 1920-2011, based on comparable metrics across countries and consistently defined over time. The methodology and data sources both have their limitations, particularly the lack of direct estimates on non-labour income, the inevitable relative narrow sectoral scope for wage dispersion, and a limited number of - though well-apportioned - occupational groups. But, given the reduced availability – and comparability – of household surveys prior to the 1980s and the difficulties of using income tax records, it offers a solid and transparent option to compare income inequality in the region, particularly over the decades of state-led protected industrialisation. The reduced number of groups is a necessity because of the difficulties in assembling a larger number of independent wage series reflecting differences in skill levels. But this has its advantages because it allows for the separation, and estimation, of high-income earners, especially property owners.

By extending the coverage up to 2011, this dataset makes it possible to bridge two contrasting periods of inequality trajectories. The first where the main sources of evidence are social tables, with the more recent one where the dominant source is official household surveys. Moreover, by facilitating comparisons with a variety of metrics it can shed light on the extent to which a common inequality story can be told regardless of the metric used. Such comparisons are also of great use to check reliability – and consistency - of my series. Trajectories of my most encompassing overall Ginis (including property income) roughly match those of alternative social-tables Ginis estimated by economic historians. Also, when a comparison is possible, movements in my overall Ginis dominated by wage income tend to be consistent with those calculated from household surveys. The use of the adopted methodology also makes it possible to separate developments within the group structure and to compare inequality dynamics in the between-group and within-group components.

A key general finding is that there is no single inequality measure that captures developments in the total occupational structure, Although, the between-group Ginis are good proxies for levels and changes in overall Ginis, the concurrence in trajectories is much weaker between overall and wage inequality. By contrast, the between-group and within-group Ginis tend to move in opposite directions. In addition, correlations across countries are both mixed in direction and strength (low in G4, high in G3W). Diversity dominates when looking at the net outcome of the underlaying forces shaping income inequality. This is puzzling; how can we make sense of this? What is behind such an outcome in countries that are similar in so many respects?

One place to look for answers is country specificities (e.g., dominant export commodities and political regimes) which can give rise to notable differences in inequality trajectories, both across countries and within the occupational structure in each country. This supports the need for more long-term country studies looking at the different inequality dimensions, where explanations of inequality patterns (including political-economy type) can be presented and discussed in depth. A second option is to explore the possibility that, despite diversity in trajectories, there is, nonetheless, a degree of commonality in the fundamentals behind them. One interpretation of the evidence presented here is that, overall, it reflects the combined action of drivers with different timings (e.g., urbanisation and demographic transition), opposite directions in their likely inequality impact (pro-labour institutions and Kuznets-type process) and varied intensity.

To disentangle their contributions, a regression analysis is required to shed light on their role at different levels of the occupational structure. I end this conclusion with an advance on the findings of a related ongoing work with Julio Revuelta (Universidad de Cantabria) that uses the new dataset to examine the role of fundamentals; moving from revealing diversity and complexity to accounting for them. The core specification is based on a model of income inequality in a small, open, developing economy built by Bourguignon and Morrison (1998).[29] We implement a dynamic five-years panel-data and use various estimation techniques to test for robustness. This econometric exercise reveals that the significance and signs of the explanatory power of the inequality drivers are contingent on the occupational Gini chosen as the dependent variable. For instance, the impact of the terms of trade (a key external variable) is disequalising in G4 and GB, but equalising in G4W and G3W, and lacks significance in G3 and G3B. When looking at policy-related drivers, a dummy variable capturing the import-substitution policy is equalising in G3 and G3B, but lacks significance elsewhere; whereas years of schooling shows a disequalising effect on within-group inequality, but lacks significance elsewhere. The only driver with explanatory power in all the six Ginis is the gap in labour productivity between manufacturing and agriculture (a proxy for the Kuznets-Lewis process). But, also, in this case there is a twist: its impact is disequalising in the overall and between-group Ginis, but equalising in the within-group component.

This new inequality dataset makes strides into the completion of the jigsaw puzzle on the long-term inequality evidence in the region, adding to previous efforts adopting various approaches and data sources. Its main virtue is to offer consistent and comparable Gini series on a sufficient number of countries and a long enough time span to inform about trajectories, commonality and diversity, not only at an aggregate level, but also at different levels of the

[29] In this model, income inequality is a function of factor endowments (labour, capital, and export-specific land and mineral resources), and their ownership structure; trade protection; foreign prices, and sectoral productivity differentials.

distributional structure. Adding pieces of evidence to the empirical puzzle allows us to have a better view of the inequality landscape that once was, and its connection with more recent, and fuller, inequality outcomes. But, at the same time, this new evidence poses new questions for our understanding of the region's inequality enigma.

## References


Allen, RC. 2019. Class structure and inequality during the industrial revolution: lessons from England's social tables, 1688–1867, Economic History Review, 72(1), 88–125.

Alvaredo F. 2010. The rich in Argentina over the twentieth century 1932-2004. In Atkinson, A. and Piketty, T. (eds.) *Top Incomes vol. II: A Global Perspective*, Oxford University Press.

Alvaredo F, Londoño Vélez J. 2013. High income and personal taxation in a developing economy: Colombia 1993-2010, CEQ wp 12.

Amarante V, Abeles M, Vega D. 2014. Participación del ingreso laboral en el ingreso total en América Latina, 1990-2010, Revista CEPAL 114 (December), 31-52.

Arroyo Abad L, Astorga P. 2017. Latin American earnings inequality in the long run. Cliometrica, 11(3), 289-295.

Arroyo Abad L, Lindert PH. 2017. Fiscal redistribution in Latin America since the nineteenth century. In Bértola L, Williamson JG (eds.) *Has Latin American Inequality Changed Direction? Looking over the Long Run*. Springer Publishing.

Astorga P. 2017a. Real wages and skill premiums in Latin America, 1900-2011. Revista de Historia Económica - Journal of Iberian and Latin American Econ. History, 35(3), 319-353.

Astorga P. 2017b. Functional inequality in Latin America: News from the twentieth century. In Bértola L, Williamson JG (eds.) *Has Latin American Inequality Changed Direction? Looking over the Long Run*. Springer Publishing.

Astorga, P. 2015. Functional inequality in Latin America: news from the twentieth century, Discussion Papers in Econ. and Social History, Oxford University, 135.

Astorga P, Bergés AR, Fitzgerald EVK. 2005. The standard of living in Latin America during the twentieth century, The Economic History Review, 59 (1), 765–796.

Azevedo JP, Inchauste G, Sanfelice V. 2013. Decomposing the recent inequality decline in Latin America, Policy Research Working Paper 6715. The World Bank.

Bacha EL. 1979. Crescimento econômico, salários urbanos e rurais: o caso do Brasil. Pesquisa e Planejamento Econômico, 5(3), 585-627.

Baptista A. 1997. *Bases Cuantitativas de la Economía Venezolana*, 1830–1995, Ediciones Fundación Polar: Caracas.

Berg EJ. 1970. Wages and employment in less developed countries. Center for Research on Economic Development, Discussion Paper 13. University of Michigan.

Bértola L. 2005. A 50 años de la curva de Kuznets: crecimiento y distribución del ingreso en Uruguay y otras economías de nuevo asentamiento desde 1870, Investigaciones en Historia Económica, 3, 135-176.

Bértola L, Williamson JG (eds.). 2017. *Has Latin American Inequality Changed Direction? Looking over the Long Run*. Springer Publishing.

Bourguignon F, Morrison C. 1998. Inequality and development: the role of dualism. Journal of Development Economics 57, 233-257

Burdín G, De Rosa M, Vigorito A, Vilá J, 2019. Was falling inequality in all Latin American countries a data-driven illusion? Income distribution and mobility patterns in Uruguay 2009-2016, dt 30/19, Instituto de Economía, Facultad de Ciencias Económicas y de Administración, Universidad de la República.

Castañeda Garza D, Bengtsson E. 2020. Income inequality in Mexico 1895-1940: Industrialization, revolution, institutions, Lund Papers in Economic History 212, Lund University, Department of Economic History.

Centro de Estudios Distributivos, Laborales y Sociales -CEDLAS, Bases de datos. Data accessed last time: January 2018.URL: http://cedlas.econo.unlp.edu.ar/
Dirección General de Estadística -DGE. 1953. Cuarto Censo Industrial de los Estados Unidos Mexicanos 1945: resumen general.
Economic Commission for Latin America and the Caribbean – ECLAC.
CEPALSTAT|Database. Data accessed last time: January 2020.
Economic Commission for Latin America and the Caribbean - ECLAC. Panorama Social in América Latina. Santiago de Chile.
FitzGerald EVK. 2008. Economic development and fluctuations in earnings inequality in the very long run: the evidence from Latin America 1900-2000, Journal of International Development 20, 1028-1048.
Flores I, Sanhueza C, Atria J, Mayer R. 2019. Top incomes in Chile: A historical perspective on Income Inequality, 1964-2017, Review of Income and Wealth, 66(4), 850-874.
Frankema EHP. 2012. Industrial wage inequality in Latin America in global perspective, 1900–2000. Stud Comp Int Dev 47, 47–74.
Frankema E. 2010. Reconstructing labor income shares in Argentina, Brazil and Mexico 1870-2000. Journal of Iberian and Latin American Economic History 28(2), 343-374.
Frankema EHP. 2009. *Has Latin America Always Been Unequal?: A Comparative Study of Asset and Income Inequality in the Long Twentieth Century*. Boston: Brill.
Gazeley I, Holmes R, Lanata Briones C, Newell AT, Reynolds K, Rufrancos H. 2018. Latin American household budget surveys 1913-1970 and what they tell us about economic inequality among households. IZA DP 11430.
Gómez León M. 2021. The Kuznets Curve in Brazil, 1850-2010, Revista de Historia Economica – Journal of Iberian and Latin American Economic History. 39(1), 37-61.
Gómez León M, Gabbuti G. 2021. Wars, depression, and fascism: Income inequality in Italy, 1900-1950, dt 2104 Universidad Pública de Navarra.
Goñi E, Lopez JH, Serven L. 2011). Fiscal redistribution and income inequality in Latin America, World Development 39,1558-1569.
Instituto Brasileiro de Geografía e Estatística -IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD).
International Labour Organization (1964-1982) Yearbook of Labour Statistics. ILO Publications: Geneva.
Jiménez JP, Gómez Sabaini JC, Podestá A. 2010. Tributación, evasión y equidad en América Latina y el Caribe, chapter 1, in: Evasión y Equidad en América Latina, CEPAL, Colección Documentos de proyectos. Santiago de Chile.
Lambert PJ, Aronson JR. 1993. Inequality decomposition and the Gini coefficient revisited. Economic Journal 103, 1221-1227.
Londoño JL. 1995. *Distribución del Ingreso y Desarrollo Económico: Colombia en el Siglo XX*, Santafé de Bogotá. Tercer Mundo Editores.
Luxemburg Income Studies Database -LIS. Data accessed last time: December 2019. https://www.lisdatacenter.org/
Maldonado LJ. 2021. Decomposing the gender pay gap in the formal sector in Venezuela: a microdata analysis 1985–2015, Applied Economics Letters, 28(14), 1145-1151.
Milanovic B, Lindert P, Williamson JG. 2010. Pre-industrial inequality, Economic Journal 121, 255–272.

Modalsli J. 2015. Inequality in the very long run: inferring inequality from data on social groups. J Econ Inequal 13, 225–247.
Morgan M, Souza P. 2019. Distributing growth during late-development: New evidence on long-run inequality in Brazil. Paper presented at CLADHE IV, July 2019, Santiago de Chile.
Pen J. 1971. *Income Distribution*, Allen Lane the Penguin Press.
Piketty, T. (2014): *Capital in the Twenty-First Century*, Harvard University Press.
Rodríguez Weber J. 2017. Nuevas estimaciones de distribución del ingreso en Colombia entre 1938 y 1988. Metodología de estimación y principales resultados. Cuadernos de Economía, Colombia, V.36 Nº 72, 43-76.
Rodríguez Weber J. 2015. The political economy of the top 1% in an age of turbulence: Chile 1913-1973, Documentos de trabajo 41, Universidad de la República.
Rodríguez Weber J. 2014, *La Economía Política de la Desigualdad de Ingreso en Chile*, Doctoral dissertation, Universidad de la República de Uruguay.
Shipley RE. 1977. On the Outside Looking In: A Social History of the Porteño Worker during the "Golden Age" of Argentine Development, 1914-1930. Ph.D. dissertation, Rutgers University.
Souza PHGF, Medeiros M. 2015. Top income shares and inequality in Brazil, 1928-2012, Journal of the Brazilian Sociological Society, 1, 119-132.
Székely M, Sámano C. 2012. Did trade openness affect income distribution in Latin America? Evidence for the years 1980–2010. UNU-WIDER Working Paper 2012-3.
Székely M, Hilgert M. 1999. *What's behind the inequality we measure: An investigation using Latin America data*, IADB wp 409.
Thorp R.1998. *Progress, Poverty and Exclusion: An Economic History of Latin America in the 20th Century*, Inter-American Development Bank publications
World Wealth and Income Database. Accessed on July 2018 from: http://wid.world/
Yap BW, Sim CH. 2011. Comparisons of various types of normality tests, Journal of Statistical Computation and Simulation, 81:12, 2141-2155.

## ANNEX A: COMPLEMENTARY TABLES AND CHARTS

TABLE A1: EMPLOYMENT SHARES BY OCCUPATIONAL CATEGORIES BY LUSTRUMS

| | Argentina | | | | Brazil | | | | Chile | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ |
| **1920** | 4.3 | 15.7 | 44.9 | 35.2 | 3.9 | 8.1 | 19.1 | 69.0 | 5.1 | 6.9 | 46.1 | 41.9 |
| **1925** | 4.3 | 16.3 | 44.2 | 35.1 | 3.8 | 7.9 | 20.4 | 67.9 | 4.7 | 6.9 | 45.9 | 42.6 |
| **1930** | 4.4 | 16.7 | 43.6 | 35.2 | 3.8 | 8.9 | 20.5 | 66.7 | 4.7 | 7.0 | 45.3 | 43.1 |
| **1935** | 4.5 | 18.2 | 42.3 | 35.0 | 3.9 | 9.6 | 21.0 | 65.5 | 5.5 | 7.1 | 45.4 | 42.0 |
| **1940** | 4.6 | 18.4 | 43.4 | 33.6 | 4.0 | 10.2 | 21.3 | 64.5 | 5.5 | 7.3 | 46.2 | 40.9 |
| **1945** | 4.8 | 18.2 | 48.7 | 28.3 | 4.0 | 10.5 | 22.5 | 63.0 | 5.4 | 7.6 | 47.8 | 39.2 |
| **1950** | 5.1 | 17.6 | 51.9 | 25.5 | 4.2 | 10.6 | 24.7 | 60.4 | 5.6 | 8.4 | 48.2 | 37.7 |
| **1955** | 5.6 | 16.5 | 54.7 | 23.2 | 5.0 | 11.2 | 26.1 | 57.7 | 6.6 | 8.7 | 49.6 | 35.0 |
| **1960** | 6.2 | 18.1 | 54.5 | 21.2 | 6.1 | 11.9 | 26.5 | 55.5 | 7.5 | 8.8 | 51.2 | 32.5 |
| **1965** | 6.4 | 19.1 | 53.9 | 20.6 | 6.9 | 11.7 | 28.9 | 52.6 | 7.9 | 10.5 | 52.9 | 28.7 |
| **1970** | 6.6 | 19.8 | 53.6 | 20.0 | 6.9 | 12.9 | 30.1 | 50.1 | 8.2 | 12.7 | 54.1 | 25.0 |
| **1975** | 7.5 | 20.4 | 53.5 | 18.6 | 6.8 | 13.6 | 32.9 | 46.7 | 10.2 | 14.1 | 51.6 | 24.1 |
| **1980** | 8.3 | 21.4 | 53.0 | 17.3 | 7.6 | 13.3 | 35.7 | 43.4 | 9.6 | 15.9 | 51.8 | 22.7 |
| **1985** | 9.0 | 23.7 | 50.4 | 16.8 | 8.0 | 13.6 | 38.1 | 40.3 | 10.6 | 14.0 | 52.8 | 22.6 |
| **1990** | 10.4 | 26.0 | 47.0 | 16.6 | 8.4 | 13.8 | 42.2 | 35.6 | 12.1 | 14.7 | 50.6 | 22.7 |
| **1995** | 11.3 | 26.2 | 46.4 | 16.2 | 7.2 | 13.4 | 45.4 | 34.1 | 12.0 | 17.1 | 48.5 | 22.4 |
| **2000** | 12.1 | 25.2 | 47.6 | 15.1 | 7.9 | 13.4 | 46.8 | 31.9 | 13.7 | 17.0 | 47.3 | 22.0 |
| **2005** | 11.4 | 25.7 | 46.4 | 16.5 | 8.1 | 13.8 | 46.8 | 31.3 | 14.4 | 17.9 | 46.4 | 21.3 |
| **2010** | 11.4 | 25.4 | 46.3 | 16.9 | 8.4 | 13.0 | 51.8 | 26.7 | 14.5 | 16.5 | 49.0 | 20.1 |
| | **Colombia** | | | | **Mexico** | | | | **Venezuela** | | | |
| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ |
| **1920** | 5.8 | 5.0 | 28.5 | 60.7 | 3.5 | 4.7 | 21.0 | 70.8 | 4.0 | 6.3 | 29.7 | 60.0 |
| **1925** | 5.7 | 5.2 | 29.6 | 59.5 | 3.4 | 5.0 | 20.8 | 70.8 | 3.8 | 6.7 | 31.7 | 57.8 |
| **1930** | 5.7 | 6.1 | 29.6 | 58.6 | 3.3 | 5.1 | 21.5 | 70.1 | 3.7 | 6.9 | 33.8 | 55.6 |
| **1935** | 6.0 | 7.7 | 28.2 | 58.2 | 3.0 | 5.5 | 23.5 | 68.0 | 3.7 | 7.3 | 35.5 | 53.4 |
| **1940** | 6.4 | 9.3 | 27.3 | 56.9 | 2.9 | 5.6 | 25.5 | 66.0 | 3.6 | 7.5 | 38.5 | 50.3 |
| **1945** | 7.0 | 9.7 | 28.7 | 54.6 | 2.9 | 5.7 | 26.3 | 65.0 | 4.3 | 7.7 | 41.8 | 46.2 |
| **1950** | 7.6 | 10.2 | 30.2 | 52.0 | 2.9 | 6.5 | 26.7 | 63.9 | 5.0 | 8.7 | 43.1 | 43.2 |
| **1955** | 7.8 | 11.6 | 30.3 | 50.3 | 3.4 | 7.1 | 29.9 | 59.6 | 5.0 | 9.7 | 44.9 | 40.4 |
| **1960** | 7.9 | 13.4 | 29.9 | 48.9 | 4.0 | 8.0 | 32.6 | 55.3 | 5.0 | 11.1 | 46.0 | 37.8 |
| **1965** | 7.7 | 14.2 | 30.8 | 47.3 | 4.9 | 8.7 | 35.9 | 50.4 | 6.0 | 13.3 | 47.6 | 33.1 |
| **1970** | 7.9 | 14.1 | 33.4 | 44.7 | 6.0 | 9.5 | 38.3 | 46.2 | 7.6 | 16.0 | 48.0 | 28.5 |
| **1975** | 8.1 | 14.6 | 34.4 | 42.9 | 6.5 | 10.3 | 37.0 | 46.3 | 9.6 | 16.9 | 45.2 | 28.4 |
| **1980** | 8.2 | 14.3 | 37.8 | 39.8 | 7.2 | 11.4 | 40.8 | 40.5 | 10.3 | 17.4 | 48.0 | 24.3 |
| **1985** | 9.6 | 14.3 | 38.8 | 37.4 | 7.7 | 12.1 | 43.0 | 37.2 | 11.2 | 16.4 | 47.1 | 25.3 |
| **1990** | 9.9 | 13.8 | 40.1 | 36.2 | 8.3 | 12.8 | 41.0 | 38.0 | 11.9 | 17.5 | 47.5 | 23.2 |
| **1995** | 11.0 | 15.5 | 37.5 | 36.1 | 8.6 | 13.3 | 41.5 | 36.7 | 11.9 | 17.9 | 46.5 | 23.7 |
| **2000** | 9.1 | 14.0 | 40.9 | 36.0 | 9.5 | 14.1 | 43.9 | 32.4 | 11.4 | 17.5 | 47.4 | 23.7 |
| **2005** | 10.1 | 13.6 | 42.8 | 33.6 | 9.9 | 14.6 | 44.7 | 30.9 | 11.5 | 16.1 | 48.1 | 24.4 |
| **2010** | 11.1 | 15.0 | 43.3 | 30.7 | 10.0 | 14.8 | 47.8 | 27.4 | 12.2 | 15.3 | 49.7 | 22.7 |

All figures in percentages (%) and three years averages. See Annex B2 for sources and estimating procedure.

TABLE A2: RELATIVE INCOME RATIOS BY OCCUPATIONAL CATEGORIES BY LUSTRUMS

| | Argentina | | | | Brazil | | | | Chile | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ |
| **1920** | 11.8 | 0.78 | 0.60 | 0.30 | 10.0 | 1.39 | 1.01 | 0.44 | 9.7 | 1.14 | 0.61 | 0.35 |
| **1925** | 10.9 | 0.86 | 0.62 | 0.31 | 10.8 | 1.13 | 0.91 | 0.46 | 11.4 | 1.20 | 0.55 | 0.31 |
| **1930** | 9.8 | 0.91 | 0.71 | 0.30 | 10.9 | 1.18 | 0.95 | 0.43 | 8.0 | 1.59 | 0.76 | 0.40 |
| **1935** | 9.4 | 0.86 | 0.74 | 0.31 | 9.6 | 1.23 | 0.88 | 0.50 | 8.8 | 1.04 | 0.66 | 0.35 |
| **1940** | 9.8 | 0.82 | 0.64 | 0.36 | 10.6 | 1.17 | 0.76 | 0.49 | 8.1 | 0.89 | 0.73 | 0.36 |
| **1945** | 10.7 | 0.76 | 0.52 | 0.32 | 12.9 | 0.97 | 0.64 | 0.42 | 7.4 | 0.98 | 0.78 | 0.38 |
| **1950** | 8.2 | 0.95 | 0.64 | 0.34 | 12.4 | 1.24 | 0.59 | 0.41 | 8.5 | 1.16 | 0.62 | 0.35 |
| **1955** | 8.1 | 0.88 | 0.60 | 0.31 | 12.8 | 1.22 | 0.59 | 0.36 | 7.8 | 0.87 | 0.65 | 0.26 |
| **1960** | 9.0 | 0.72 | 0.48 | 0.24 | 13.1 | 1.26 | 0.61 | 0.29 | 5.6 | 1.08 | 0.73 | 0.35 |
| **1965** | 8.3 | 0.78 | 0.50 | 0.27 | 13.3 | 1.13 | 0.56 | 0.30 | 5.1 | 0.96 | 0.76 | 0.33 |
| **1970** | 7.8 | 0.79 | 0.51 | 0.30 | 12.0 | 0.97 | 0.66 | 0.26 | 5.9 | 1.03 | 0.59 | 0.24 |
| **1975** | 7.1 | 0.70 | 0.52 | 0.24 | 11.1 | 0.86 | 0.53 | 0.23 | 5.5 | 0.70 | 0.59 | 0.16 |
| **1980** | 7.0 | 0.67 | 0.48 | 0.16 | 9.2 | 0.89 | 0.53 | 0.22 | 5.7 | 0.71 | 0.58 | 0.17 |
| **1985** | 5.7 | 0.76 | 0.52 | 0.26 | 7.6 | 0.95 | 0.66 | 0.20 | 4.9 | 0.91 | 0.55 | 0.24 |
| **1990** | 5.5 | 0.69 | 0.45 | 0.23 | 7.1 | 1.10 | 0.69 | 0.21 | 4.8 | 0.75 | 0.53 | 0.21 |
| **1995** | 5.5 | 0.67 | 0.38 | 0.19 | 6.4 | 1.20 | 0.69 | 0.20 | 5.1 | 0.68 | 0.48 | 0.19 |
| **2000** | 4.9 | 0.73 | 0.40 | 0.21 | 6.5 | 1.13 | 0.58 | 0.20 | 4.4 | 0.70 | 0.50 | 0.19 |
| **2005** | 4.7 | 0.81 | 0.46 | 0.22 | 5.9 | 1.10 | 0.62 | 0.25 | 4.2 | 0.71 | 0.49 | 0.19 |
| **2010** | 4.2 | 0.89 | 0.54 | 0.26 | 5.5 | 1.09 | 0.62 | 0.29 | 4.3 | 0.68 | 0.47 | 0.18 |

| | Colombia | | | | Mexico | | | | Venezuela | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ |
| **1920** | 6.1 | 1.52 | 1.11 | 0.41 | 13.8 | 1.35 | 0.65 | 0.45 | 10.6 | 1.48 | 0.94 | 0.35 |
| **1925** | 7.3 | 1.29 | 0.92 | 0.38 | 12.5 | 1.39 | 0.75 | 0.49 | 9.9 | 1.56 | 1.01 | 0.35 |
| **1930** | 6.1 | 1.58 | 1.19 | 0.34 | 11.9 | 1.72 | 0.78 | 0.51 | 8.3 | 2.09 | 1.03 | 0.37 |
| **1935** | 7.7 | 1.38 | 0.93 | 0.34 | 11.3 | 1.96 | 0.78 | 0.54 | 8.7 | 1.63 | 0.98 | 0.40 |
| **1940** | 8.0 | 1.21 | 0.92 | 0.33 | 11.5 | 1.66 | 0.96 | 0.49 | 8.1 | 1.60 | 1.07 | 0.36 |
| **1945** | 7.6 | 1.22 | 0.92 | 0.33 | 18.1 | 1.26 | 0.71 | 0.32 | 7.1 | 1.38 | 1.07 | 0.32 |
| **1950** | 7.4 | 1.17 | 0.82 | 0.33 | 19.1 | 1.09 | 0.70 | 0.28 | 8.9 | 1.15 | 0.82 | 0.25 |
| **1955** | 7.5 | 1.13 | 0.72 | 0.30 | 15.8 | 0.96 | 0.72 | 0.30 | 8.6 | 1.17 | 0.78 | 0.27 |
| **1960** | 6.4 | 1.29 | 0.78 | 0.31 | 11.9 | 0.96 | 0.77 | 0.35 | 10.2 | 1.04 | 0.60 | 0.27 |
| **1965** | 6.3 | 1.18 | 0.74 | 0.28 | 8.5 | 1.06 | 0.82 | 0.38 | 7.6 | 1.17 | 0.61 | 0.30 |
| **1970** | 6.2 | 1.18 | 0.68 | 0.28 | 8.6 | 0.89 | 0.64 | 0.34 | 5.3 | 1.13 | 0.67 | 0.36 |
| **1975** | 6.8 | 0.90 | 0.63 | 0.29 | 6.2 | 1.01 | 0.83 | 0.40 | 5.5 | 0.97 | 0.51 | 0.32 |
| **1980** | 6.0 | 1.04 | 0.69 | 0.35 | 7.0 | 0.90 | 0.59 | 0.38 | 5.7 | 0.91 | 0.40 | 0.29 |
| **1985** | 5.4 | 1.09 | 0.72 | 0.34 | 6.7 | 0.87 | 0.59 | 0.34 | 5.8 | 0.79 | 0.36 | 0.24 |
| **1990** | 5.6 | 0.98 | 0.66 | 0.34 | 6.2 | 0.96 | 0.60 | 0.32 | 5.8 | 0.62 | 0.34 | 0.18 |
| **1995** | 5.9 | 0.97 | 0.64 | 0.33 | 5.4 | 1.09 | 0.69 | 0.29 | 5.7 | 0.59 | 0.37 | 0.25 |
| **2000** | 5.6 | 0.94 | 0.61 | 0.31 | 5.3 | 0.98 | 0.62 | 0.26 | 4.5 | 0.97 | 0.53 | 0.30 |
| **2005** | 5.4 | 0.91 | 0.55 | 0.30 | 5.2 | 0.92 | 0.59 | 0.28 | 4.9 | 0.80 | 0.51 | 0.24 |
| **2010** | 5.0 | 0.87 | 0.53 | 0.28 | 5.4 | 0.84 | 0.55 | 0.27 | 5.0 | 0.72 | 0.48 | 0.18 |

All figures are three years averages, except Mexico in c.1920 which excludes 1919. Income ratios are calculated using estimated household income per person engaged in the denominator. See Annex B2 for sources and estimating procedure.

## TABLE A3: NORMALITY TESTS

| | **Tests: Shapiro-Wilk, Anderson-Darling, Lilliefors, and Jarque-Bera** | | |
|---|---|---|---|
| | Non rejection of H0 at the 5% significance level | | rejection of H0 & acceptance of Ha at 5% level |
| | *in four or three tests* | *in two tests* | *in four or three tests* |
| **Industrial surveys** | | | |
| *blue-collar workers* | Ar1917(13), Ar1937(63), Ar1963(19); Br1920/28(70), Br1949(20), Br1973/84(18), Br1984(21); Ch1928(20), Ch1937/57/67 (22), Ch1953(19), Ch1975/80/**87**(27); Co1934(33), Co1936(20), Co1942(26), Co1963(19), Co1976/86(27); **Mx1940**(50), Mx1946/47/49(32), Mx1960(19); Ve1953(17), Ve1971/76(26) | **Br1959**(20); Mx1950(32), **Mx1990**(26); **Ve1986**(25) | **Mx1930**(44), **Mx1948**(32)**,** **Mx1986**(29) |
| *while-collar workers* | Ar1963(19); Br1959(20), Br1973(21), Br1984(21); Ch1928/37(18), Ch1953/57/67 (22), Ch1980/84(27); Co1936(20), **Co1942**(26), Co1963(19), Co1976/86(27); Mx1960(19), Mx1986/90(26); Ve1953(17), Ve1971/76(26), Ve1986(25) | | **Ch1975**(27) |
| **Occupational surveys** | | | |
| *low-skilled occupations* | Mx1935/36(14) rural & urban | | |
| *low-skilled urban occupations ILO's OI* | Ar1936(8) | | **Ch1938**(8) |
| *blue-collar & construction workers (ILO'OI)* | Co1938(17); Ve1943(20) | **Mx1940**(17) | **Ar1936**(20); **Ch1938**(20) |
| *semi-skilled workers* | Mx1935/36(25) urban | | |
| **Household budget surveys** | | | |
| *c1-c35* | Br1976/1995; Ch1992/1998/2009; Co2007/2010; Mx1984/92/98/2004 | **Ve1993/2005** | **Br1985/2005; Ve1985** |
| *c36-c70* | **Br1985/**1995/2005; Ch1992/1998/2009; Co2007/2010; Mx1984/92/98/2004; Ve1993 | | **Br1976; Ve1985/2005** |
| *c71-c90* | Br1976/1985/1995/2005; Ch1992/1998/2009; Co2007/2010; Mx1984/92/98/2004; Ve1985/1993/2005 | | |
| *c71-c95* | **Mx1984**; **Ve1993/**2005 | **Br1976/1985/1995; Ch1992/1998/2009; Co2007/2010; Mx1992/1998/2004; Ve1985** | **Br2005** |

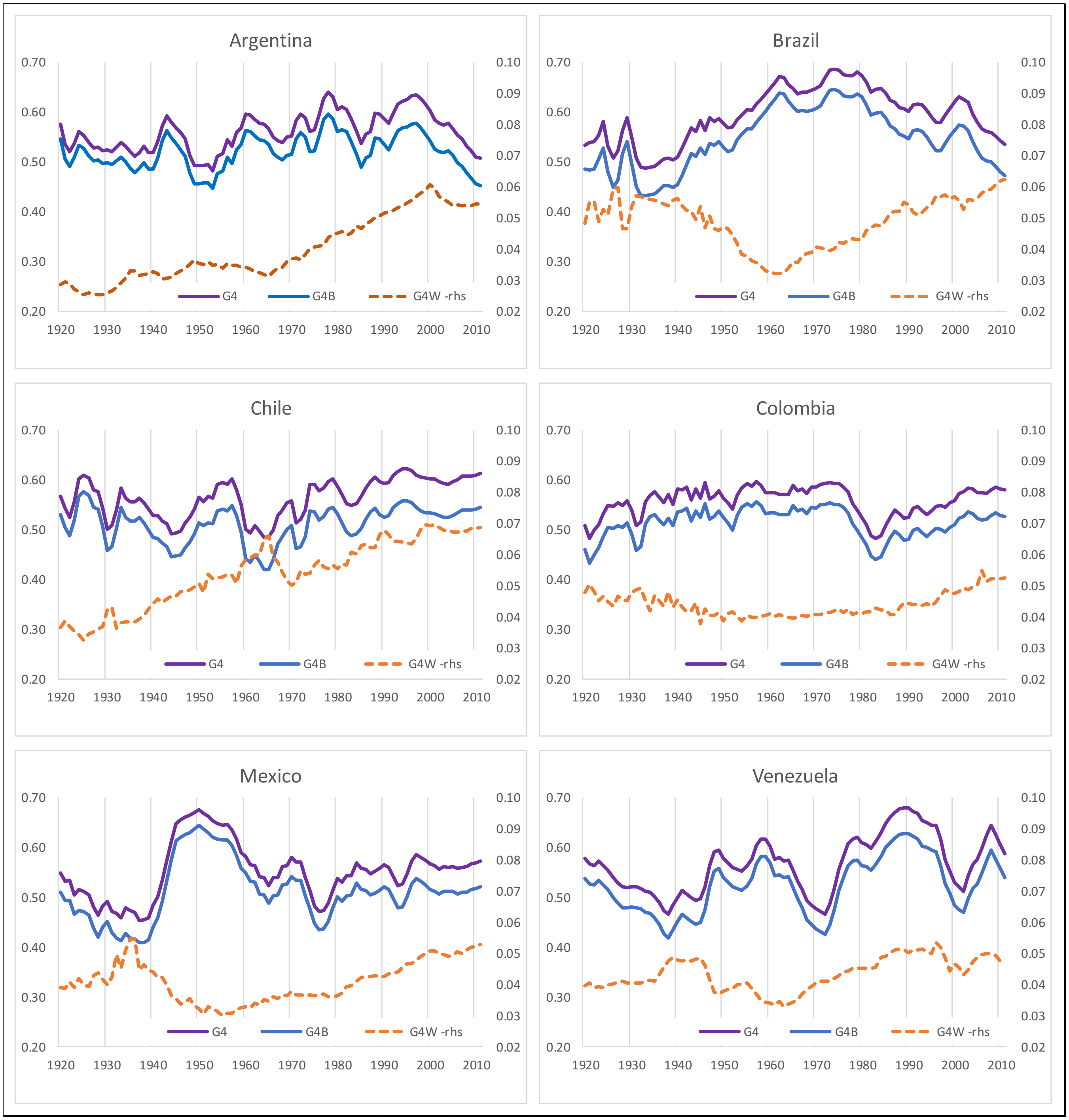
H0: The variable from which the sample was extracted follows a Normal distribution. Ha: The variable from which the sample was extracted does not follow a Normal distribution. Figures in brackets stand for the number of observations. Observations in bold: the Shapiro-Wilk test is rejected. See Annex C for sources.

TABLE A4: INCOME OVERLAPS BETWEEN THE THREE LOWER GROUPS, SELECTED YEARS

| | Group 4 | | | | | Group 3 | | | | | Group 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $-2\sigma_4$ | $-1\sigma_4$ | $u_4$ | $+1\sigma_4$ | $+2\sigma_4$ | $-2\sigma_3$ | $-1\sigma_3$ | $u_3$ | $+1\sigma_3$ | $+2\sigma_3$ | $-2\sigma_2$ | $-1\sigma_2$ | $u_2$ | $+1\sigma_2$ | $+2\sigma_2$ | $u_1/u_2$ |
| **Argentina** | | | | | | | | | | | | | | | | |
| **1920** | 19 | 25 | 30 | 36 | 41 | 43 | 52 | 60 | 69 | 78 | 59 | 72 | 84 | 97 | 110 | 14.1 |
| **1940** | 38 | 46 | 54 | 62 | 69 | 57 | 77 | 96 | 115 | 135 | 82 | 102 | 123 | 143 | 163 | 11.9 |
| **1960** | 43 | 49 | 54 | 60 | 65 | 74 | 91 | 108 | 125 | 141 | 107 | 134 | 160 | 186 | 213 | 12.6 |
| **1980** | 31 | 35 | 39 | 43 | 47 | 78 | 105 | 132 | 159 | 185 | 110 | 148 | 186 | 225 | 263 | 12.1 |
| **2000** | 17 | 32 | 46 | 61 | 76 | 48 | 69 | 91 | 112 | 133 | 76 | 120 | 164 | 208 | 252 | 6.7 |
| **Brazil** | | | | | | | | | | | | | | | | |
| **1920** | 9 | 14 | 20 | 26 | 31 | 26 | 36 | 46 | 56 | 66 | 37 | 50 | 63 | 76 | 89 | 7.4 |
| **1940** | 10 | 18 | 26 | 34 | 42 | 24 | 32 | 41 | 49 | 58 | 38 | 50 | 63 | 75 | 88 | 9.1 |
| **1960** | 13 | 20 | 26 | 33 | 40 | 38 | 47 | 56 | 65 | 74 | 74 | 95 | 115 | 136 | 156 | 10.5 |
| **1980** | 13 | 27 | 41 | 54 | 68 | 45 | 71 | 97 | 123 | 148 | 103 | 135 | 168 | 201 | 233 | 10.5 |
| **2000** | 13 | 25 | 36 | 48 | 59 | 36 | 70 | 105 | 139 | 174 | 80 | 142 | 203 | 265 | 327 | 5.8 |
| **Chile** | | | | | | | | | | | | | | | | |
| 1920 | 15 | 20 | 25 | 30 | 34 | 27 | 35 | 43 | 51 | 59 | 50 | 65 | 80 | 95 | 110 | 8.5 |
| 1940 | 17 | 23 | 30 | 36 | 42 | 33 | 46 | 60 | 73 | 86 | 44 | 59 | 73 | 87 | 101 | 9.2 |
| 1960 | 18 | 28 | 38 | 47 | 57 | 35 | 56 | 78 | 99 | 121 | 68 | 92 | 116 | 140 | 164 | 5.1 |
| 1980 | 19 | 25 | 31 | 37 | 43 | 60 | 90 | 121 | 151 | 181 | 79 | 106 | 133 | 160 | 187 | 9.3 |
| 2000 | 27 | 45 | 64 | 82 | 100 | 62 | 115 | 168 | 221 | 274 | 134 | 184 | 234 | 283 | 333 | 6.3 |
| **Colombia** | | | | | | | | | | | | | | | | |
| **1920** | 6 | 11 | 15 | 19 | 23 | 20 | 29 | 38 | 48 | 57 | 23 | 38 | 53 | 68 | 83 | 4.1 |
| **1940** | 11 | 17 | 24 | 30 | 36 | 31 | 48 | 66 | 83 | 100 | 41 | 64 | 86 | 109 | 132 | 6.6 |
| **1960** | 12 | 18 | 25 | 31 | 38 | 36 | 49 | 62 | 74 | 87 | 55 | 80 | 106 | 131 | 156 | 5.0 |
| **1980** | 22 | 32 | 42 | 51 | 61 | 47 | 66 | 85 | 104 | 124 | 73 | 101 | 128 | 156 | 184 | 5.7 |
| **2000** | 22 | 33 | 43 | 54 | 65 | 42 | 63 | 84 | 105 | 126 | 68 | 99 | 130 | 161 | 192 | 5.9 |
| **Mexico** | | | | | | | | | | | | | | | | |
| **1920** | 19 | 24 | 30 | 35 | 40 | 23 | 31 | 38 | 45 | 53 | 45 | 62 | 79 | 96 | 114 | 11.5 |
| **1940** | 24 | 33 | 42 | 51 | 60 | 50 | 66 | 82 | 97 | 113 | 73 | 107 | 141 | 175 | 210 | 7.0 |
| **1960** | 28 | 37 | 45 | 53 | 62 | 65 | 83 | 100 | 117 | 134 | 74 | 100 | 125 | 151 | 176 | 12.3 |
| **1980** | 66 | 80 | 94 | 109 | 123 | 102 | 130 | 157 | 184 | 212 | 153 | 191 | 229 | 266 | 304 | 6.9 |
| **2000** | 28 | 40 | 51 | 62 | 73 | 61 | 90 | 119 | 148 | 176 | 98 | 144 | 189 | 234 | 279 | 5.4 |
| **Venezuela** | | | | | | | | | | | | | | | | |
| **1920** | 8 | 11 | 15 | 18 | 21 | 24 | 32 | 39 | 47 | 55 | 31 | 46 | 62 | 78 | 94 | 7.2 |
| **1940** | 13 | 20 | 27 | 35 | 42 | 44 | 63 | 82 | 101 | 121 | 62 | 92 | 123 | 153 | 183 | 5.1 |
| **1960** | 26 | 38 | 50 | 62 | 74 | 75 | 93 | 110 | 128 | 145 | 122 | 156 | 191 | 226 | 260 | 9.8 |
| **1980** | 63 | 79 | 95 | 110 | 126 | 89 | 112 | 135 | 158 | 181 | 201 | 250 | 299 | 348 | 398 | 6.5 |
| **2000** | 27 | 40 | 52 | 65 | 77 | 61 | 76 | 90 | 105 | 119 | 107 | 137 | 167 | 196 | 226 | 4.7 |

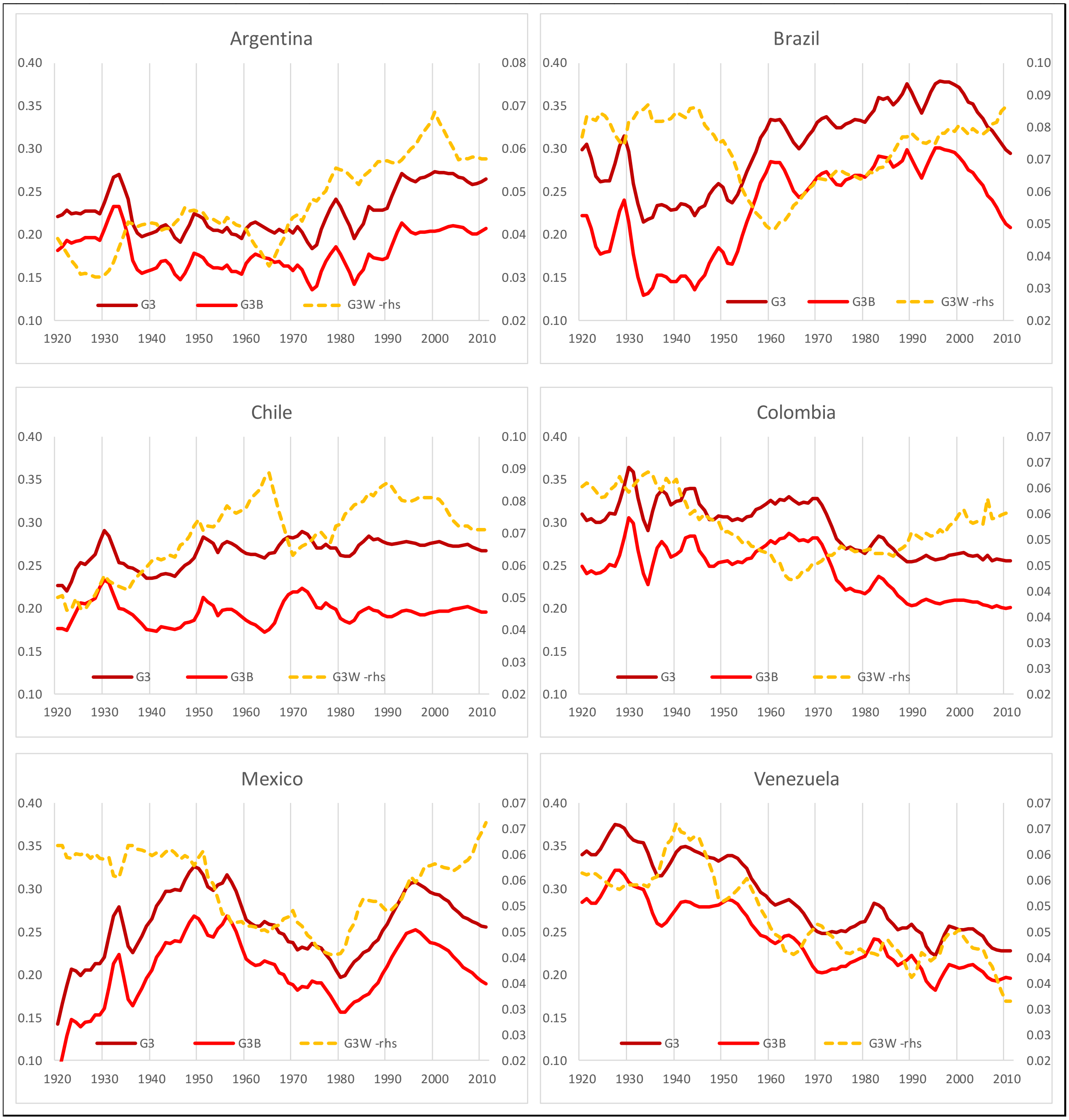
Own calculations based on the “68-95-99.7 rule”. All figures in percentages are three-year centred averages.

FIGURE A1: OCCUPATIONAL GINIS WITH FOUR GROUPS BY COUNTRY




All three-years moving averages. G4W is plotted on the right-hand side.

FIGURE A2: OCCUPATIONAL GINIS WITH THREE WAGE GROUPS BY COUNTRY




All three-years moving averages. G3W is plotted on the right-hand side.

## ANNEX B: ESTIMATION PROCEDURE AND DATA SOURCES

### *B1: Income shares*

To calculate the three income shares, I need to make a series of adjustments to the underlying income and wage data.

1. *Common price deflators*

First, I need to use the same price deflator in both series of overall income and real wages. The original GDP series at 1970 prices are multiplied by the ratio of the GDP implicit deflator to the CPI. In this way both GDP and real wage series are deflated by the CPI (see Astorga, 2015).

2. *From GDP to household income*

According to my estimation methodology (*see Section 2 of the main paper*), if the series of GDP at market prices are used as a proxy for household income levels, it will result on a significant overestimation of the income share of the top 10% of the labour force and an underestimation of the share of the remaining 90%. This is so because items such as indirect and corporate taxes, repatriated profits, the consumption of fixed capital and the net surplus of the public sector will be allocated to the income share of the occupational top group (which is the main component of the top 10% of the EAP).[30] Therefore, the GDP series need to be adjusted downwards so as to reflect household income.

To do such an adjustment, when available, I use national account data of Household Income (HI). A ratio of HI to GDP (with both variables at current prices) is calculated and, then, applied to the original series of GDP per EAP at 1970 prices to make the required level adjustment. When the HI/GDP ratio is not available, I use changes in alternative ratios to move backward and, in some cases, forward from the last HI/GDP datapoint. These alternative ratios are calculated using Private Consumption (on the household income outlays side) or more encompassing income concepts such as Private Income (Private Consumption plus Private Savings), or National Income. Otherwise indicated, all series are sourced from ECLAC website, CEPAL's Boletín Económico de América Latina (BEAL,1961 and 1962), ECLA's Statistical Bulletins (SBLA, 1961-1972), and ECLAC's Statistical Yearbooks (SYLA) from 1973 onwards.

Estimation details by country are as follows:

- Argentina: HI data in 1951-1961, 1965, 1967-1973. The years 1962 to 1964 and 1966 are interpolated based on changes in private consumption. Between 1947 and 1951 the ratio HI/GDP growths in line with the ratio of private consumption to GDP (BCRA,1976), and between 1935 and1950 with that of national income to GDP (BCRA,1976). Between 1935 and 1951 I use National Income, and prior to 1935 I use the growth rates of the GDP series at 1970 prices.

[30] I am not making allowances for realized capital gains. These are a significant source of income at the top in developed economies as many corporations distribute profits using share repurchases instead of dividends (Atkinson et al., 2011, note in p.35). But this is likely to be less significant in the LA-6 during most of the period. In any case, the exclusion of capital gains would result in an underestimation of the income share of the top 10%.

- Brazil: HI data in 1947-1960 (FGV, *Contas Nacionais do Brasil*) and in 2000-2009 (ECLAC). Estimates of National Income in 1939 and 1947 (ECLA, SBLA 1972) are used to extend the ratio back to 1939. And prior to 1939 I use the growth rates of the GDP series at 1970 prices. The gap between 1961 and 1969 is filled with linear interpolation. Between 1970 and 1999 I use the ratio of private income to GDP. This is estimated based on data on the share of private savings on GDP (IBGE website) and an assumed capital depreciation of 5% of Gross National Income. After 2009 I use the ratio of private consumption to GDP.
- Chile: HI data in 1958-1976, 1996-2011. During the 1976-1996 period I use the ratio of Private Income to GDP. Between 1940 and 1958 I use the ratio of Private Consumption to GDP (both in real terms from Díaz et al. 2016). Prior to 1940 I use a proxy for national income estimated by Javier Rodriguéz Weber by deducting profits from foreign mining companies (copper and nitrates) from GDP figures in Haindl (2007).
- Colombia: HI in 1950-2011. Between 1900 and 1950 I use the ratio of private consumption to GDP (De Corso, 2019).
- Mexico: HI in 1993-2011. Between 1950-1960 I use the ratio of Private Consumption to GDP; and the ratio of National Disposable Income to GDP in 1960, 1965, 1970-77 (ECLAC's SYB), and 1980-1993 (INEGI *Cuentas Nacionales*). The gaps are filled with liner interpolation. Prior to 1960 I use the growth rates of the GDP series at 1970 prices.
- Venezuela: HI in 1950-1959 (including retained profits, from CEPAL's BEAL 1961, statistical appendix), 1960-1969 and 1978-2007. I use the ratio of Private Consumption to GDP to fill the gaps in 1900-1950 (using De Corso, 2013),1970-1977 and 2007-2011.

3. *Reconciling wage data with the national accounts:*

In the national accounts overall income is divided into several functional categories: employment (including wages and other labour income), profits, rents, and self-employment (a mix of property and labour income). In order to calculate the relative income levels for each of the occupational groups consistent with the national accounts, I proceed as follows:

First, I calculate the wage income in c.2000 estimated for the three lower occupational categories (WI234). A bulk of the self-employed in my sample of countries are likely to belong to the informal sector and to be part of the two lower occupational categories. Their mixed income (largely labour income) is assumed to be equal to the average wage of the corresponding group. At the other end of the self-employment spectrum, the income of the owner-managers or self-employed professionals are estimated as part of the residual (see Chapter 2 in the paper).

Secondly, I use estimates of national accounts' labour income c.2000 (including the labour income component in mixed income) for each of the LA-6 countries in Amarante et al. (2014, *Cuadro 2*, *Estimación* 2).[31] From this total, I need to deduct the component accruing to professionals which is not part of my three lower occupational groups. This is done by using three pieces of information: ratios of incomes of professionals relative to workers in commerce; the share of professionals in the EAP c.2000 (both in ECLAC, 2000); and the monthly earnings of workers in the sector of retailing and commerce (ILO website). The resulting aggregate for labour income excluding that of professionals is LINA234.

[31] The estimated labour income as share of GDP at factor cost are: 45.7% for Argentina in 2000, 56.6% for Brazil in 2001, 54.9% for Chile in 2000, 52.8% for Colombia in 2009, 45.6% for Mexico in 2000, and 46.5% for Venezuela in 2000.

The next step is to calculate the share of LINA234/GDP and to divide my WI234 by this ratio in order to obtain a level of GDP c.2000 (GDP*) which is consistent with the proportionality between labour income and GDP in the national accounts. This procedure is also intended to capture some of the fringe payments that are largely excluded from my wage data.

Finally, I need to adjust downwards GDP* in c.2000 to reflect household income (see previous section); and call this HI*. The resulting estimated share of WI234/HI* c. 2000 are: 46% for Argentina, 53% for Brazil, 46.3% for Chile, 51.6% for Colombia, 53.3% for Mexico, 51.5 for Venezuela. Then, starting from this 2000 HI* benchmark, I use the growth rate of the estimated household income per worker (at 1970 prices) series to go back to 1900 and forth to 2011.

### *B2: Economically active population shares*

I followed two different procedures to estimate the shares of the EAP per each of the four occupational categories:

<u>Period 1950-2011</u>

The employment shares of the four groups are estimated by aggregating categories for the distribution of the EAP by occupational groups according to data collected by the International Labour Organization (ILO) and ECLAC. I use four different classifications:

1. ECLAC (2000). Group 1: employers, managers and professionals. Group 2: technicians and administrators (clerks). Group 3: urban workers (retailing and transport, excluding low skilled workers and street vendors), artisans and blue-collar workers. Group 4: rural workers and personal services (includes domestic servants) plus low skilled urban workers and street vendors.
2. ILO, ISCO-88. Group 1: 1 legislators, senior officials and managers; 2 professionals Group 2: 3 technicians and associate professionals; 4 clerks; plus 6 skilled agricultural and fishery workers. Group 3: 5 service workers and shop and market sales workers; 7 craft and related trade workers; 8 plant and machine operators and assemblers; 0 armed forces; Group 4: 9 elementary occupations; plus X not classifiable by occupation.
3. ILO, ISCO-68. Group 1: 0/1 professional, technical and related workers; 2 administrative and managerial workers. Group 2: 3 clerical and related workers; (1/2)* 4 sales workers. Group 3: (1/2)* 4 sales workers; 7/8/9 production and related workers, transport equipment operators and labourers; X not classifiable by occupation. Group 4: 5 Service workers; 6 agriculture, animal husbandry and forestry workers, fishermen and hunters
4. PREALC (1982). Group 1: 0 professional, technical and related workers; 1 managerial workers. Group 2: 2 clerical and related workers; (1/2)*3 sales workers. Group 3: (1/2)*3 sales workers; 5 transport; 6-7 artisans and blue collar workers. Group 4: 4 agriculture; 8 service workers.

I use simple interpolation to fill the gaps in each of the occupational structures calculated with the above sources. I take the ECLAC figures for circa 2000 to set the share levels and then go backwards using information on changes in each of the four categories provided by the additional three classification systems (in ISCO-88, ISCO-68 and PREALC). To splice series from two different occupational structures I use a common year and then apply rate of changes to go backwards. The data available in each of the classifications by country are:

*Argentina*. ISCO-88 in 1998-2006. PREALC (1982) in 1960, 1970. Interpolations: 1960-70. 2000-06: uses ISCO-88 with a correction for the methodological break in 2003.

*Brazil*. ISCO-88 in 2000, 2002-07. ISCO-68 in 1981-90; 1992-93; 1995-99; 2001, 1971, and 1983 (in ILO Yearbooks) and 1976-2006 available online. PREALC (1982) in 1950, 1960, 1970. Interpolations: 1951-59; 1961-69; 1991; 1994; 2000. 2000-07: it uses ISCO-88. Shares in 2000 and 2001 are as in 2002.

*Chile*. ISCO-88 in 2002. ISCO-68 in 1960, 1971, and 1983 (in ILO's Yearbooks), and 1976-2006 available online. The categories "mining" and "armed forces" are included in Group 3. PREALC (1982) 1952, 1960, 1970. In 1952 the total of categories 0 to 3 are split according to the structure in 1960. Interpolations: 1953-59; 1961-70; 1972-75. 2000-08: it uses ISCO-68.

*Colombia*. ISCO-68 in 1975-80; 1985-87; 1989-90; 1992-2000; 2001-08. Data exclude the armed forces and are based on surveys on seven main cities. PREALC (1982) and ILO's Yearbooks in 1951, 1964, 1973 (only ILO), and 1980. In 1951 the categories "managerial workers" and "clerical and related workers" are split according to the structure in 1964. Interpolations: 1952-63; 1965-72; 1974-79. 2000-08: it uses ISCO-68. Shares in 2000 are as in 2001.

*Mexico*. ISCO-88 in 2000. ISCO-68 in 1988; 1991; 1993; 1995-2008. PREALC (1982) and ILO (Yearbooks, compatible with ISCO-68) in 1950, 1960, 1970, 1975 (ILO), 1980 (ILO). Interpolations: 1951-59; 1961-69; 1971-74; 1976-79; 1981-87.2000-08: it uses ISCO-88.

*Venezuela*. ISCO-68 in 1976-2008. PREALC (1982) and ILO (Yearbooks, compatible with ISCO-68) in 1950, 1961, 1971 (ILO), 1981 (ILO). In 1950 the categories "professionals", "managerial workers" and "clerical and related workers" are split according to the structure in 1961. Interpolations: 1951-60; 1962-70; 1972-80. 2000-08: it uses ISCO-68.

In all six countries for the final years of the current century the estimation is as follows: shares of Group 1 are kept equal to the last data point; those of Group 2 are estimated based on changes in the EAP share of manufacturing (ECLAC); for Group 4 I use changes in the share of agriculture; shares for Group 3 are estimated as a residual.

Period 1920-1950

To complete the employment shares back from 1950 to 1900 I rely on changes in three indicators constructed by FitzGerald (2008) as follows: Group 1, the stock of university graduates as a proportion of the total of those with primary education. The stock of educational graduates is found using the perpetual inventory method applied to the data on enrolment in primary and tertiary education. Group 2, total employment in manufacturing and public administration as a proportion of the EAP. Manufacturing employment comes from census data, and public administration employment is estimated from levels of government expenditure. Group 3 is estimated as the residual from the other three groups. Group 4, the agricultural share of the EAP, from census data. This includes not only agricultural workers as such, but also small farmers (i.e., peasants) and family labour on a non-wage basis.

### *B3: Additional data sources*

CPI series are as in Astorga (2012), except Chile 1920-1940 from Haindl (2007). In Argentina to avoid the under-reporting of CPI inflation by INDEC in 2006-11, I use a CPI index reported by seven provinces compiled by CENDA.

*GDP at constant 1970 prices*: Chile 1920-1940, I deflate the nominal GDP series in Haindl (2007) with the CPI from the same source. Mexico 1900-1910 uses Estadísticas Económicas del Porfiriato, available at the ITAM website. In Venezuela I use De Corso (2013). Otherwise, I use MOxLAD.

*GDP's Implicit deflators*: Argentina 1920-2004 from Ferreres (2005); IBGE website for Brazil 1900-2011; Chile 1920-1970 from Haindl (2007) and MOxLAD 1970 to 2000; Colombia 1920-1996 from GRECO and MOxLAD thereafter; Mexico from MOxLAD 1921 to 2000; Venezuela 1920-2011 from De Corso (2013).

*Wage series***:** see Astorga (2017a, online Appendix). For Mexico I updated the wage series for unskilled workers between 1983 and 2000. It is now based on manufacturing average wage in industries 311-322 from ILOSTAT, which are dominated by relatively unskilled labour.

## References for Annex B


Amarante V, Abeles M, Vega D. 2014. Participación del ingreso laboral en el ingreso total en América Latina, 1990-2010, Revista CEPAL 114 (December), 31-52.

Astorga P. 2017a. Real wages and skill premiums in Latin America, 1900-2011. Revista de Historia Económica - Journal of Iberian and Latin American Econ. History, 35(3), 319-353.

Astorga P. 2012. Mean reversion in long-horizon real exchange rates: evidence from Latin America. Journal of International Money and Finance 31(6): 1529-1550.

Atkinson A, Piketty T, Saez E. 2011. Top incomes in the long run of history, Journal of Economic Literature 49(1), 3-71.

Banco Central de la República Argentina (BCRA). 1976. *Series Históricas de Cuentas Nacionales de la Argentina*. Publicaciones BCRA.

Centro de Estudios para el Desarrollo Argentino (CENDA). CPI for seven Provinces from January 2005 to December 2011. Accessed in February 2013.

De Corso G. 2019. Crecimiento económico colombiano de 1888 a 2013: una nueva serie del producto interno bruto, Revista de Economía Institucional 21(40), 259-289.

De Corso G. 2013. El crecimiento económico de Venezuela desde la oligarquía conservadora hasta la revolución bolivariana: 1830-2012. Una visión cuantitativa, Revista de Historia Económica 31(3), 321-358.

Díaz JR, Lüders R, Wagner G. 2016. Chile 1820-2000: *La República en Cifras*, Ediciones Universidad Católica de Chile, Santiago, Chile

Economic Commission for Latin America and the Caribbean (ECLAC): CEPALSTAT|Database. Data accessed: January, 2014.

–ECLAC. 2000. *Panorama Social in América Latina*. Santiago de Chile.

–ECLAC. 1973-2001. *Statistical Yearbook for Latin America* (SYLA). UN, New York, NY.

Economic Commission for Latin America (ECLA). 1964-1972. Statistical Bulletin for Latin America (SBLA). UN, New York, NY.

Comisión Económica para América Latina (CEPAL). 1961, 1962. Boletín Económico de América Latina, Suplemento Estadístico. Naciones Unidas, Santiago de Chile.

Ferreres O. 2005. *Dos Siglos de Economía Argentina (1810-2004)*. Fundación Norte y Sur.

FitzGerald EVK. 2008. Economic development and fluctuations in earnings inequality in the very long run: the evidence from Latin America 1900-2000, Journal of International Development 20, 1028-1048.

Fundação Getulio Vargas (FGV), *Contas Nacionais do Brasil, 1947-1980*. FGV Publications, Rio de Janeiro.
Grupo de Estudios del Crecimiento Económico (GRECO). 2002. *El Crecimiento Económico Colombiano en el Siglo XX*. Banco de la República - Fondo de Cultura Económica, Bogotá.
Haindl E. 2007. *Chile y su Desarrollo Económico en el Siglo XX*. Editorial Andrés Bello, Santiago de Chile
International Labour Organization Databank (ILOSTAT). Last accessed on February 2021 https://ilostat.ilo.org/
International Labour Organization. 1964-1982. *Yearbook of Labour Statistics*. ILO Publications: Geneva.
International Labour Organization. 1936-1964. International Labour Review. ILO Publications: Geneva.
Instituto Brasileiro de Geografía e Estadística (IBGE) Databank. Accessed on January 2015 http://www.ibge.gov.br/
Instituto Brasileiro de Geografía e Estadística (IBGE). Contas Nacionais do Brasil, 1986-2000. http://www.ibge.gov.br/
Instituto Nacional de Estadística Geografía e Informática (INEGI). various years. Sistema de Cuentas Nacionales de México. Mexico City, DF: INEGI.
Instituto Nacional de Estadística Geografía e Informática (INEGI). 1990. *Estadísticas Históricas de México*. Mexico City, DF: INEGI.
Montevideo-Oxford Latin American Economic History Database (MOXLAD). The Latin American Centre of Oxford University, and Universidad de la República (Uruguay).
Programa Regional de Empleo para América Latina y el Caribe (PREALC).1982. *Mercado de Trabajo en Cifras*. 1950-1980. PREALC publications.

## ANNEX C: WAGE DISPERSION OF THREE LABOUR CATEGORIES IN LATIN AMERICA: 1920-2011

This annex describes the procedure adopted, the sources used and the assumptions made in constructing time series for wage dispersion for unskilled workers, as well as for blue- and white-collar workers I manufacturing during the 1920-2011 period for Argentina, Brazil, Chile, Colombia, Mexico and Venezuela. In all cases wage dispersion is measured by the coefficient of variation (*cv*), and is defined in a consistent and comparable way across the six countries.

I largely use wages on rural and urban occupations (e.g., labourers) when estimating dispersion of unskilled wages; whereas, in the case of blue and white-collar workers, I largely rely on data from industrial censuses or surveys with a breakdown by industries (e.g., food, footwear etc.). There is also data on blue-collar occupations (e.g., from the International Labor Organization October Inquiry – ILO/OI) since the end of the 1930s or early 1940s. However, a mayor disadvantage of this source is a patchy coverage, undermining comparability. However, when a comparison is possible, trends in wage dispersion for blue-collar workers by occupations and industries tend to move in line.[32]

More specifically, for unskilled workers I rely on country censuses and surveys complemented by daily/hourly wages in a selection of unskilled occupations in manufacturing, construction and services from the ILO/OI. In some cases, and for the more recent decades, I use income dispersion at the lower section of the distribution as reported in official household budget surveys (HBS). A key issue in this group of workers is the rural-urban divide, as I am covering a period in which the region underwent a rapid process of internal migration. To capture its impact on wage dispersion, when data allows, I assemble a representative sample of unskilled wages for both rural and urban activities in benchmark years, with the proportions in the sample matching the corresponding urbanisation rate at the time. In this way it is possible to account for wage disparities within both sectors, as well as between them in Brazil and Colombia. In Mexico, owing to limited data on rural wage dispersion, the series incudes within-urban dispersion and the rural-urban wage gap. In Argentina and Chile, where the urbanisation rate was already over 40% in the 1920s (Astorga et al., 2005), wage dispersion is driven by the urban sector. This is also the case in Venezuela where rural wage data are scant.

For blue- and white-collar workers, the main data source is official industrial surveys in manufacturing following the International Standard Industrial Classification (ISIC) breakdown by industries (divisions). The data up to the 1980s usually refer to ISIC1 (up to 20 industries), and to ISIC2 (up to 28 industries) thereafter. The comparison of *cv* values across industries shows that discrepancies between ISIC1 and ISIC2 tend to be small (in many cases within a

---

[32] In Chile during the 1937-69 period the average *cv* of a group of up to 20 occupations from ILO/OI is 30.1% (calculated based on 16 yearly observations); whereas the average *cv* for blue-worker wages in manufacturing (ISIC1 of up to 20 industries) is 31.3% (based on 6 yearly observations). Also, there are matching dispersion trends over periods with coinciding observations: a rise between 1937-1953, a fall from 1953 to 1957, a rise from 1957 to 1963, and a fall from there to 1968. More generally, Modalsli (2015) reports that for the recent decades in Latin America - using microdata - the dispersion of the occupational structure is similar to that of industry.

+/- 5% interval).[33] Of particularly important are *cv* discrepancies between ISIC1 and ISIC2 in years in which there is a change from the first to the second version. Some examples are: Argentina in 1971 for blue-collar wages (ISIC2=19.8% vs. ISIC1=19.5%), and in 1974 for white-collar wages (ISIC2=21.7% vs. ISIC1=20.4%); Chile in 1973 for blue-collar wages (23.5% vs. 23.8%), and in 1972 for white-collar wages (17.6% vs.16.7%); Mexico in 1975 for blue-collar wages (19.4% vs.19.3%); and in 1985 for white-collar wages (21.9% vs. 19.4%). Wages in petroleum & coal activities (including refineries) are excluded from the calculations. This is a highly capital-intensive industry with a significant rent component that translates in unusually high wages and salaries relative to other manufacturing industries.

When constructing the series of wage dispersion, if necessary, two interpolations procedures are employed to fill the gap between two given data points $y_0$ and $y_n$: linear interpolation, and pattern interpolation which uses information of a known auxiliary series *x* to fill the gap in *y*.[34] Unless otherwise indicated, linear interpolation is the default procedure. Finally, for each country a continuous annual series of wage dispersion in each of the three categories are constructed by smoothing the available data points at five-year intervals.

In the following by-country descriptions, the sub-periods that include the year in which the dispersion level is set are preceded by an asterisk. Starting from that year the series of coefficient of variation move backwards and forwards using changes in related dispersion series.

## Argentina

*Unskilled workers* ($cv_{unsk}$):
*1920-1935: dispersion in daily wages for unskilled men in 11 industries in the city of Buenos Aires from Shepley (1977, Table IX-A). Data points in 1917, 1921-22, 1926, 1929. Wage dispersion in 1935 is assumed to equal that of blue-collar workers in the city of Buenos Aires (see below).
1935-1960: it grows with dispersion in hourly wages (*salarios básicos mínimos de convenio*) for unskilled workers in 14 occupations (covering industry and services activities) in the city of Buenos Aires. Data points in 1934-40, 43, 46, 49, 51, 54, 57 and 1960 from Cuesta and Newland (2017).
1960-1996: unskilled wage dispersion in manufacturing (national level) from ILO YLS, with data points in 1960-75 (ISIC1); 1976-1994 (ISCI2) except in 1978 and 1980. Between 1994 and 1996 it uses wage dispersion of all blue workers (see below).
1996-2011: the *cv* of mean earnings of deciles 1 to 3 from World Bank database (based on official household surveys) in 1996, 1999, 2002, 2005, 2008 and 2011.

---

[33] Frankema (2011) also found that long-term trends hardly differed from the ISIC1 and ISIC2 classifications – also with ISIC3.

[34] For a given year "t" in the interpolated period $y_0$-$y_n$, the in-between values are estimated according to the following expression: $y_t = y_{t-1}$ * [($x_t$ / $x_{t-1}$)]/ [($x_n$/$x_0$)/($y_n$/$y_0$)]1/n. This method is used in Willebald (2011) and Rodríguez Weber (2014).

*Blue-collar workers* ($cv_{b-c}$):
1920-1935: dispersion in daily wages for blue-collar male workers in 11 industries in the city of Buenos Aires from Shipley (1977, Table IX-A), calculated as weighted averages of unskilled and relatively skilled workers using Shipley's weights. Data points in 1917, 1921-22, 1926, 1929. The figure in 1920 is assumed equal to that in 1917. Wage dispersion in 1935 is calculated for blue-collar wages in 11 industries (matching those included in 1929) in the city of Buenos Aires using the second industrial census (Dirección General de Estadística de la Nación, 1937). This data point is used to splice the series with the one starting in 1935 based on national data.
*1935-1985: industrial censuses at the national level in 1935, 1947 (ISIC1, data of 1946), 1954 (data of 1953), 1963, 1974 (ISIC2), and 1985. Industrial surveys in 1938 (ISIC1), 40, 42, 48, 1950-52, 1954-62 and 1964 from ILO YLS; and in 1971 (ISIC2), 1976, 1979, and 1982 (ISIC2) from UN Yearbook of Industrial Statistics - YIS.
1985-2011: blue-collar wage dispersion from ILO YLS (ISIC2 to 1994, and ISIC3 thereafter).

*White-collar workers* ($cv_{b-c}$):
*1920-1935: the dispersion level is set in 1935 using census data for salaries of white-collar workers (Dirección General de Estadística de la Nación, 1937) at a national level. To go back to 1920, dispersion grows in line with that of blue-collar workers (see above).
1935-1993: industrial censuses in 1935 (ISIC1), 1947 (data of 1946), 1954 (data of 1953), 1963, 1974 (ISIC2), 1985. Industrial surveys in 1971 (ISIC2), 1976, 1979, and 1982 (ISIC2) from UN YIS. In 1989 and 1993 estimates are based on changes in skilled-wage dispersion from ILO YLS (1992, 1996).
1993-2011: dispersion of earnings per employee in overall manufacturing in 1993 (ISIC2), 1996, 1998 to 2001 from UN YIS, and in 2002 to 2011 (ISIC3) from Ministerio de Trabajo, Empleo y Seguridad Social website.

When not indicated, the sources for industrial surveys and censuses are United Nations The Growth of World Industry, and United Nations Yearbook of Industrial Statistics.

## Brazil

*Unskilled workers* ($cv_{unsk}$):
*1920-1959: dispersion is estimated based on three comparable benchmarks which take into account the proportional representation of rural and urban unskilled workers.

1. In 1920 the coefficient of variation is calculated using male daily unskilled wages (national averages) for a total of fourteen occupations: ten rural (*arador, carreiro, carroceiro, trabalhador de enxada hombre, cortador de cana, derribador madera, lenhador, odenhador, vaqueiro, oleiro* - all wages "*sem sustento*"); and four urban (*fiandeiro, cardador, tintureiro* (textile industry) and *acabador* (shoe industry). These four occupations are taken as representatives of relatively low skilled urban jobs. The proportions of rural and urban occupations match the urbanisation rate in 1920 (27.4% in the population census). All sourced from the 1920 census (*Diretoria Geral de Estatística* – DGE, 1928, V).
2. In 1936 the calculation is based on a similar number and composition of rural and urban occupations as in 1920. The estimated urbanisation rate in 1936 is 30.1%, as in the population census of 1940 (IBEG, 1949). The four representative urban unskilled wages

are for workers in Rio de Janeiro and Sao Paulo in construction, electricity, printing, and mechanical engineering, sourced from ILO/OI (ILR, 1937). These wages are scaled up to the national level by using the ratio of the corresponding average wages in the two cities to the ten main cities in Brazil using data for 1943 (ILR, 1945). The 1936 estimation is extended to 1939 (payroll data year of the second industrial survey, IBGE 1949) by following two steps. First, constructing a national series of rural unskilled wages for ten occupations in 21 states using as weights the states' shares of economically active population in agriculture in 1940 in IBGE's Anuario Estatístico do Brasil (AEB 1941-45, p.30). Data are available for 1937 and 1938. The dispersion in 1939 is estimated assuming the same rate of growth between 1937 and 1938. Secondly, estimating wages for the same four urban occupations based on data on ten main cities in 1943 (ILO ILR, 1945). The 1939 value is obtained as the simple average of wages in 1936 and 1943. Two additional data points are estimated in 1924 and 1934 using changes in dispersion in rural wages (national aggregates calculated for three occupations: *arador, trabalhador de enxada* and *tratador animais*) available in 1920 and 1924, and in 1934 and 1936 from IBGE AEB.

3. In 1959 (the wage data year for of the fourth industrial and agricultural censuses, IBGE 1963 & 1967 respectively) I use national averages of daily unskilled wages for a total of sixteen occupations: nine rural (*arador, carreiro, enxada, cortador de cana, lenhador, tratador animals, vaqueiro, oleiro*, and the average of the preceding eight occupations - all wages without meals); and seven urban (construction, printing, mechanical engineering, textile, chemicals, steel, and the average of the six preceding occupations) from ILO/OI (ILR, 1959). The latter data are for hourly wages in October 1958. Daily values are calculated assuming 8 hours per day, and then extended to 1959 using the growth rate of blue-collar median wages in the respective industries (AEB, 1962, p.195). The proportions of rural and urban occupations reflect the 45% urbanisation rate in 1960 (IBGE, 1963). To fill the gap between 1939 and 1959 I use pattern interpolation using as auxiliary series blue-collar wage dispersion in seven industries dominated by low-skilled labour (ISIC1, divisions 20, 21, 23, 24, 25, 26, and 29) with data points in 1939, 47, 49, 52, 55, 58, and 1959 from IBGE's AEB (various years). The figure in 1944 is an interpolation.

1959-1981: dispersion of blue-collar wages in seven relatively low-wage industries (IBGE AEB, various years).
1981-2011: earnings dispersion in centiles 1 to 35 (excluding zero income observations) from IBGE PNAD (Brazil's household budget surveys) in years 1981, 81, 85, 89, 92, 95, 99, 2001, 2005, 2009 and 2011.

*Blue-collar workers* ($cv_{b-c}$):
*1920-1949: in 1920 uses wage dispersion in 12 industries from the 1920 population census (DGE, 1928, vol. V). A data point in 1928 is estimated using industrials surveys for the Federal District (9 industries) in 1920 and 1928 from the same source. The 1924 figure is interpolated. Observations in 1939 and 1949 are based on wage dispersion in 22 industries from the second and third industrial censuses (IBGE 1949, 1957). Data points in 1936, 1941 and 1943 are estimated using changes in wage dispersion in ILO/OI for 11 blue-collar occupations; in 1936 and 1943 it uses wages in Rio de Janeiro and Sao Paulo (ILR, 1937 and 1945), and in 1941 of the Federal District (ILR, 1942). The figure in 1939 is interpolated.
1949-2011: official industrial surveys or censuses in 1949, 52, 55, 58, 59, 62, 66, 69, 73, 74, 76, 77, 79, 81, 84, 85, 88, and 1992 (ISIC1 from IBGE website), plus 1956, 57, 64, 65, and

1993-95 (ISIC2) from ILO YLS. Observations in 1996, 1999, 2001, 2004, 2007, 2010 are for ISIC3 from IBGE website.

*White-collar workers* (*$cv_{b-c}$*):
*1920-1949: Observations of while-collar wage dispersion in 1937 from IBGE AEB (1938, p.341) for 19 industries (ISIC1), and in 1939 from AEB (1947). The series is extended backwards to 1928 and 1920 using changes in dispersion series for blue-collar wages. Values in 1933 and 1944 are interpolated.
1949-2011: white-collar wages in official industrial censuses or surveys in 1949, 52, 55, 58, 59, 62, 66, 69, 73, 74, 76, 77, 79, 81, 84, 85, 88, and 1992; all ISIC1 from IBGE website. Data points in 1996, 1999, 2001, 2004, 2007, 2010 are for ISIC3 from IBGE website.

## Chile

*Unskilled workers* (*$cv_{unsk}$*):
*1920-29: wage dispersion in eight low-skills occupations (6 in industry, 1 in mining, 1 in rural areas) from Rodriguez Weber (2014). Data points in 1920, 23, 25, 29.
1929-1986: ILO/OI (ILR, various years) average hourly wages in Santiago de Chile of a number of unskilled occupations in manufacturing, utilities, construction and public services: 8 occupations in 1936, 39, 41, 44, 46, 50, 1951; 10 occupations from 1951 to 1965 except 1954 and 1961; 19 occupations in 1985 and 1986. To splice the series with the observation in 1929, I used the ratio of wage dispersion of four similar occupations in both sources (construction, furniture, metals work, and day labourer). To fill the gap from 1965 and 1985 I applied pattern interpolation using as the auxiliary series wage dispersion in industries dominated by relatively unskilled workers (ISIC1, divisions 20-26) from ILO YLS (1965, 1972, 1978, 1980). From the interpolated series I took 3-year averages in 1968, 71, 74, 77, 80, and 1983.
1986-2008: earnings dispersion in centiles 1 to 30 from HBS in 1986, 92, 94, 96, 98, 2000, 2006, and 2009 from the Luxembourg Income Study Database (LIS).

*Blue-collar workers* (*$cv_{b-c}$*):
1918-1928: real wage dispersion of blue-collar workers in 17 industries from Matus (2009).
*1928-67: Rodriguez Weber (2014) wage dispersion of blue-collar workers in 1928, 1937, 1953, 1957, 1963 and 1967. Between 1937 and 1953 I apply pattern interpolation using as the auxiliary series hourly wage dispersion in 20 occupations (excluding unskilled ones) from ILO/OI (ILR, various years) in 1937, 38, 40, 41, 44, 46, 47, 50, 51, and 1953. Changes in this series is also used to calculated dispersion in 1958, 60 and 1967.
1967-1980: wage dispersion from United Nations (1973) and UN YIS; ISIC1 (20 industries) up to 1972, ISIC2 (28 industries) up to 1980.
1980-1993: wage dispersion in 28 ISIC2 industries from ILO YLS.
1993-2009: earnings dispersion in the percentiles 31 to 70 in 1986, 92, 94, 96, 98, 2000, 2006, 2009 from LIS.

*White collar workers* (*$cv_{w-c}$*):
1918-1928: real wage dispersion of white-collar workers in 17 industries from Matus (2009).
*1928-67: wages dispersion in white-collar workers in 1928, 1937, 1957 and 1967 from Rodriguez Weber (2014); and in 1953 and 1963 from UN (1953-1965). Value in 1960 is

interpolated. Intermediate estimates in 1941, 1945 and 1949 are based on a white-collar Gini series (Rodriguez Weber, 2014).
1967-1984: UN (1973) and UN YIS (1975-82); ISIC1 up to 1971, ISIC2 up to 1984.
1984-2005: wage dispersion of all manufacturing workers (ISIC2) in 1985, 89, 92, 96, 98, 2000, 2003, and 2005 from UN YIS and UNIDO International Yearbook of Industrial Statistics. Dispersion from 2005 to 2011 is assumed to equal that in c.2005.

## Colombia

*Unskilled workers* ($cv_{unsk}$):
*1920-1936: wage dispersion in 1936 is calculated by combining wages in five urban unskilled occupations (*vendedores ambulantes, cobradores de buses, albañiles, pintores*, and *latoneros*) from Dirección General de Estadística - DNE Anales de Economía y Estadística (1936), with 10 rural labourers' daily wages (without meals) in 10 states (*departamentos*) from DNE Anuario General de Estadística -AGE (1942, p.216). Because limited unskilled wage data on rural occupations, dispersion across states is used as a proxy. The rural and urban weights are in line with the urbanisation rate of 29.1% in 1938 (AGE, 1946, p.vi). Wage dispersion in 1920, 1925, and 1930 are assumed equal to that in 1936.
1936-1988: uses income dispersion calculated from lower levels of the decile structure of rural labourers (d1-d4) and of urban employees (d1-d3) in 1938, 1952, 1964, 1971, 1978, and 1988 from Londoño (1995). Each benchmark includes a combination of ten rural and urban mean incomes; with the proportions of each category matching the urbanisation rate in the respective year. When necessary, I calculated additional values in between deciles from Londoño's original estimates. This procedure captures both the between rural-urban gap and the within dispersion in both groups.
The series starting in 1938 is spliced with the 1936 benchmark by using changes in the dispersion of daily rural wage (without meals) across 17 states between both years – which implicitly assumes a constant mean and dispersion in the urban component – sourced from Romero et al. (2000). Intermediate values in 1944 and 1957 are estimated by following a similar procedure. The value in 1983 is interpolated.
1988-2011: dispersion grows in line with the income ratio of unskilled workers with basic schooling (up to five years) to those without schooling in 1988, 1991, 1994, 1997, 2000, 2003, and 2005 (all circa values) from Dirección Nacional de Planeación´s Estadísticas Históricas de Colombia. Value in 2008 equals that of 2005.

*Blue-collar workers* ($cv_{b\text{-}c}$):
1920-1936: starting from the 1936 datapoint the *cv* grows backwards with changes in the dispersion of average daily wages of five blue-collar occupations in the Fenicia factory from Urrutia and Arrubla (1970). Three-years centered averages are calculated every three years. Figure in 1923 is interpolated.
*1936-1945: wage dispersion in 1936 (22 industries), 1938 (25 industries), 1939, 1941 and 1942 from DNE AGE. The breakdown of these data is not fully compatible with that in the first industrial census of 1945 (DNE, 1947). To splice the series, the 1945 value is assumed equal to that in 1942.
1945-2011: official industrial censuses or surveys in 1945 (ISIC1), 1953, 1963, 1967, 1970 (ISIC2), 1976, 1986, 1992, and 1996 from Departamento Adminstrativo de Estadística

(DANE).[35] Figures in 1992, 2002, and 2007 are calculated based on changes in wage dispersion for permanent workers only (DANE website). Between 1945 and 1953 the series grows in line with dispersion in blue-collar wages in 11 industries (ISIC1) in ILO YLS (various years); and between 1956 and 1963 it grows with dispersion in 15 industries in DANE AGE (various years) available at DANE website. Intermediate values in 1971-1975 and 1977-1980 are calculated based on the dispersion of annual earnings at 1970 prices (DANE website).

*White-collar workers* ($cv_{w-c}$):
1920-1936: starting from 1936 the *cv* grows backwards to 1920 with changes in the dispersion of average monthly salaries of seven white-collar occupations in the public sector from López Uribe (2008). Three-years centered averages are calculated every three years.
*1936-1945: I use white-collar wage dispersion of up to 25 industries in 1936, 1938, 1939, 1941, and 1942 from DNE's AGE. The breakdown of these data is not fully compatible to that in the first industrial census of 1945 (DNE, 1947) adopting ISIC1. To splice the series, the figure in 1945 is assumed equal to 1942.
1945-2011: official industrial censuses or surveys in 1945 (ISIC1), 1949 (interpolated), 1953, 1963, and 1967, 1970 (ISIC2), 1976, 1986, 1992, 1996; and 2002 & 2007 (based on permanent workers only). Between 1956 and 1963 and in 1968 the series grows in line with dispersion in 15 industries in DANE AGE (various years). Intermediate values in 1971-1975 and 1977-1980 are calculated based on the dispersion of annual earnings at 1970 prices available at DANE website.

## Mexico

*Unskilled workers* ($cv_{unsk}$):
(*)1920-1940: the dispersion level is set in the years of 1935 and 1936 based on data of 20 low-skilled occupations available in the Dirección General de Estadística (Anuario Estadístico de México-AEM, 1938, pp.146-51); of which 10 are rural and 4 are urban. The proportions reflect the estimated urbanisation rate of 35% circa 1940 (MOxLAD). From the 1935-36 benchmark, the dispersion series grows forward to 1940 with yearly changes in the hourly-wage dispersion of 14 activities plus the rural minimum wage (see below). The series are extended back to 1929 by using changes in the yearly wage dispersion over a similar group of activities in 1929 and c.1940 from Castañeda and Bengtsson (2020). And from 1929 to 1923 by using wage dispersion in low-skilled industries (pottery, leather, textiles, clothing, construction) from Departamento de la Estadística Nacional-DEN (1930, p.88) plus the minimum agriculture wage (INEGI, EHM, vol I, p.182). This captures the rural-urban wage gap in low skilled workers plus within-urban dispersion. Values from 1920 to 1922 equal wage dispersion circa 1925.
1940-1977: wage dispersion is calculated using hourly wages of industrial activities available from 1934 to 1977 in the Anuarios Estadísticos de México (AEM) on yearly basis except in the years of 1961 and 1962. I selected 14 activities of relatively low hourly wages, namely: vegetable oils, cigars, footwear, carpentry, cement, tanning, wax products, cotton processing, sweets and chocolate, building materials, milling(wheat), biscuits, canning, and working cloths.

[35] The Departamento de Administrativo General de Estadística replaced the Dirección General de Estadística in 1953 as Colombia´s statistics agency responsible for the Anuario General de Estadística and other censuses and surveys.

To account for changes in rural wages, I used data on the rural official minimum daily wage at the national level available during the same period on biannual basis in the AEM. In order to minimise any adjustment lag to inflation in the second year of each biennial, I took the first data point in each pair of the rural minimum wage and interpolate the second year with the first data point of the following biannual pair. The resulting wage series (converted to hourly wage by assuming 8 hours in a working day) is added to the 14 urban wage series and the coefficient of variation calculated. In this way, the estimated dispersion also captures changes in the rural-urban gap in low skilled wages. As not sufficient data are available to account for rural dispersion across occupations or activities, it is assumed that it remains stable during the period.
1977-1992: it uses the coefficient of variation of mean wages in 9 manufacturing industries (divisions 311-332) dominated by relatively low-skilled workers.
1992-2011: it grows in line with the income dispersion in the percentiles 1 to 35 in the HBS (non-zero income) from INEGI website. Biannual data.

*Blue-collar workers* ($cv_{b\text{-}c}$):
(*)1920-1940: in 1929 and c.1940 wage dispersion in calculated using matching 40 blue-collar activities from Castañeda and Bengtsson (2020). The data point in 1934 is calculated using blue-collar wage dispersion in 14 matching manufacturing industries in 1930 (I industrial census, data from 1929) and 1935 (II industrial census, data from 1934) in Dirección General de Estadística AEM (1953). The series are extended back to 1924 using the daily-wage dispersion in 8 manufacturing industries (food, textiles, clothing, leather, furniture, pottery, construction materials, and metal products) from DEN (1930). Values from 1920 to 1923 are assumed to equal the dispersion in c.1925.
1940-2011: to estimate the intermediate values between 1940 (III industrial census, DGE 1953) and 1960 (VII industrial census, DGE 1965-ISIC1) I used pattern interpolation based on yearly hourly-wage dispersion across 25 manufacturing activities from AEM (various years). Going forward, the 1963 data point is calculated based on changes in hourly-wage dispersion in 25 manufacturing activities. The data point in 1965 is calculated using industrial data from D.F. and Estado de Mexico (ISIC1), accounting for half of total employment in manufacturing in that year. Then, industrial surveys for the whole country are available in 1968, 1971, 1975 (ISIC2), 1978, 1981, 1985, 1988, 1990, 1992, and 1994 (INEGI Biblioteca Digital). In 1999, 2004 and 2009 the series grows in line with wage dispersion based on ISIC2 available at INEGI website.

*White-collar workers* ($cv_{w\text{-}c}$):
(*)1920-45: white-collar wage dispersion across 14 matching manufacturing industries (ISIC1) in 1929, 1934, 1940, 1945 are from the I, II, III and IV industrial censuses (INEGI, 1953). Values in1920 and 1925 are assumed to equal the dispersion in 1930.
1945-1960: wage dispersion in overall manufacturing in 1945 and 1950 (V industrial census, DGE 1957) and in 1960 (VII industrial census, DGE 1965). Value in 1955 is an interpolation.
1960-211: white-collar wage dispersion from industrial censuses in 1960 (ISIC1), 1955 (interpolated), 1965 (data from D.F. and Estado de Mexico), 1971 (industrial survey), 1975, 1980, 1985 (ISIC2), 1988, 1990 (industrial survey), and 1994 (INEGI Biblioteca Digital). In 1999, 2004, and 2009 white-collar wage dispersion (ISIC2) from INEGI website.

## Venezuela

*Unskilled workers* ($cv_{unsk}$):
1920-1940: wage dispersion grows from 1940 back to 1936 using changes in wage dispersion in blue-collar workers (see below). Values in 1920, 1925, and 1930 are assumed equal to that in 1936.
1940-1990: Average hourly wages for men in eight unskilled occupations (construction, electricity, city councils, printing, mechanical engineering, conductors, goods porter and per-way labourers) in Caracas from ILO/OI (ILR) and ILO YLS. Data available for 1940, 1943, 1946, 1950 and 52 (6 occupations), 1957 (interpolated), 1962 (7 occupations including textiles, chemicals and steel), and in 1965, 1966, 1971 (10 occupations). In 1976 and 1980 dispersion grows in line with that of blue-collar wages. In 1984, 1986, 1989, 1990 wage dispersion across up to 20 relatively unskilled occupations (daily averages, men only) from ILO LABORSTA.
1990-2011: dispersion in the income centiles 1 to 30 (non-zero labour income per person) calculated from official household budget surveys (Maldonado, 2021).

*Blue-collar workers* ($cv_{b-c}$):
1920-1936: Dispersion figures in 1920, 25, 30, 35 assume the same value as in 1936.
1936-1986: data points in 1936 (first industrial census, from Valecillos, 1990, p.27), 1953 and 1971 from UN (1953-1965) - ISIC1; and in 1986 from UN YIS (1990) – ISIC2. Between 1941 and 1953 the series grows in line with the dispersion across 20 hourly wages of adult males in up to 20 occupations (excluding unskilled labour) from ILO/OI (ILR, 1936-1964). Between 1965 and 1971, and from1971 to 1984 it grows in line with the *cv* of average monthly earnings of blue-collar workers from ILO YLS (ISIC1 to 1976 and ISIC2 thereafter).
1986-2011: it grows in line with the income dispersion in centiles 31 to 70 (non-zero labour income per person) calculated from official HBS (Maldonado, 2021).

*White-collar workers* ($cv_{w-c}$):
1920-1936: wage dispersion in 1920, 1925, 1930, 1935 assume the same value as in 1936.
1936-1986: data points in 1936 (first industrial census, from Valecillos 1990), 1944 (interpolated), 1953 (ISIC1) and1971 (ISIC2) from UN (1953-1965), 1976; and 1986 from UN YIS (1990). Values in 1961 (ISIC1), 1966 and 1981 (ISIC2) are based on changes in wage dispersion in all manufacturing from Valecillos (1990).
1986-2011: income dispersion in centiles 71 to 90 (non-zero labour income per person) calculated from official HBS (Maldonado, 2021).

### References for Annex C


Baptista A. 1997. *Bases Cuantitativas de la Economía Venezolana, 1830–1995*. Ediciones Fundación Polar, Caracas.

Castañeda Garza D, Bengtsson E. 2020. Income inequality in Mexico 1895-1940: Industrialization, revolution, institutions, Lund Papers in Economic History 212, Lund University, Department of Economic History.

Cuesta EM, Newland C. 2017. Peronismo y salarios reales. Otra mirada al período 1939-56. Investigaciones y Ensayos, Academia Nacional de la Historia, Buenos Aires, (Jan-June), 64, 75-98.

Departamento Administrativo Nacional de Estadística -DANE. 1953. Anuario General de Estadística Colombia 1951-52, Bogotá, Imprenta Nacional.
Dirección Nacional de Estadística -DNE. 1936. Anales de Economía y Estadística 1936, vol 1(1). Contraloría General de la República. Bogotá, Colombia.
–DNE. 1942. Anuario General de Estadística, Colombia 1941, Bogotá, Publicaciones de la Contraloría General, Imprenta Nacional.
–DNE. 1946. Anuario General de Estadística, Colombia 1945, Bogotá, Imprenta Nacional.
–DNE. 1947. Primer Censo Industrial-1945, Contraloría General de la República. Bogotá, Imprenta Nacional.
Dirección Nacional de Planeación -DNP. Estadísticas Históricas de Colombia, capítulo 6 – sector laboral. DNP webpage. https://www.dnp.gov.co/estudios-y-publicaciones/estudios-economicos/Paginas/estadisticas-historicas-de-colombia.aspx
Departamento de la Estadística Nacional. 1930. Anuario de 1930, 2da época, 16. Estados Unidos Mexicanos.
Dirección General de Estadística -DGE.1938-1977. Anuarios Estadísticos de los Estados Unidos Mexicanos - AEM). Talleres Gráficos de la Nación, Mexico D.F.
–DGE.1952. Tercer Censo Industrial de los Estados Unidos Mexicanos-1940: resumen general. Talleres Gráficos de la Nación, Mexico D.F.
–DGE. 1953. Cuarto Censo Industrial de los Estados Unidos Mexicanos-1945: resumen general. Talleres Gráficos de la Nación, Mexico D.F.
–DGE. 1957. Quinto Censo Industrial de los Estados Unidos Mexicanos-1950: resumen general. Talleres Gráficos de la Nación, Mexico D.F.
– DGE. 1965. Séptimo Censo Industrial 1961 de los Estados Unidos Mexicanos: resumen general. Talleres Gráficos de la Nación, Mexico D.F.
Dirección General de Estadística de la Nación. 1937. Segundo Censo Industrial - Argentina
Diretoria Geral de Estatística -DGE. 1928. Recensamento Geral do Brasil 1920 (vol. V). Rio de Janeiro: tipografía de estatística.
Frankema EHP. 2012. Industrial wage inequality in Latin America in global perspective, 1900–2000. Stud Comp Int Dev 47:47–74.
Instituto Brasileiro de Geografía e Estatística -IBGE webpage. http://www.ibge.gov.br/
–IBGE. 1990. *Estatísticas Históricas do Brasil: Séries Econômicas, Demográficas e Sociais de 1550 a 1988*. Second edition, IBGE, Rio de Janeiro.
–IBGE. 2003. *Estatísticas do Século XX*. Rio de Janeiro: IBGE. In: http://www.ibge.gov.br/seculoxx/default.shtm
–IBGE. 1924, 1934-1939, 1941/1945-1954. Anuario Estatístico do Brasil. Rio de Janeiro.
–IBGE.1945 onwards. Anuario Estatístico do Brasil. Rio de Janeiro: typhografía da estatística.
–IBGE. 1949. Recensamento Geral do Brasil 1940. Rio de Janeiro: typhografía da estatística.
–IBGE. 1957. Recensamento Geral do Brasil 1950. Rio de Janeiro: typhografía da estatística.
–IBGE. 1963. Recensamento Geral do Brasil 1960. Rio de Janeiro: typhografía da estatística.
–IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD.
Instituto Nacional de Estadística y Censos de la República Argentina -INDEC. http://www.indec.mecon.ar/
Instituto Nacional de Estadística Geografía e Informática -INEGI. 1990. *Estadísticas Históricas de México*. Mexico City, D.F.
Instituto Nacional de Estadística Geografía e Informática -INEGI, Biblioteca Digital https://www.inegi.org.mx/app/publicaciones/default.html?t=0730001000000000&ag=99

Instituto Nacional de Estadísticas de Chile -INE. http://www.ine.cl/
International Labour Organization. 1936-1964. International Labour Review. ILO Publications: Geneva.
International Labour Organization. 1964-1982. *Yearbook of Labour Statistics*. ILO Publications: Geneva.
International Labour Organization, LABORSTA website. http://laborsta.ilo.org
Londoño JL. 1995. *Distribución del Ingreso y Desarrollo Económico. Colombia en el Siglo XX*, Bogotá, Tercer Mundo Editores, Banco de la República, Fedesarrollo.
Lopéz Uribe MP. 2008. Diferenciación salarial y condiciones de vida en Bogotá, 1900-1950, *Documentos CEDE* 25, November.
Luxembourg Income Study Database -LIS. https://www.lisdatacenter.org/our-data/lis-database/
Matus M. 2009. *Precios y Salarios Reales en Chile durante el Ciclo Salitrero, 1880-1930*. Ph.D. dissertation. Universitat de Barcelona, Barcelona.
Ministerio del Trabajo, Empleo y Seguridad Social de Argentina. http://www.trabajo.gov.ar/
Montevideo-Oxford Latin American Economic History Database, MOXLAD. The Latin American Centre, Oxford University, and the Universidad de la República (Uruguay). http://www.lac.ox.ac.uk/moxlad-database
Rodriguez Weber J. 2014. *La Economía Política de la Desigualdad del Ingreso en Chile, 1850-2009*. Doctoral Thesis, Universidad de la República, Uruguay.
Romero CA, Jaramillo CF, Nupia O. 2000. Integración en el mercado laboral colombiano: 1945-1998. Borradores de Economía,148, Banco de la República.
Shipley RE. 1977. *On the Outside Looking In: A Social History of the Porteño Worker during the "Golden Age" of Argentine Development, 1914-1930*. Ph.D. dissertation, Rutgers University.
Urrutia M, Arrubla M (ed.). 1970. *Estadísticas Históricas de Colombia*, Universidad Nacional de Colombia, Bogotá.
United Nations -UN. 1953-1965. *The Growth of World Industry*, National Tables, New York.
United Nations. 1973. *The Growth of World Industry*, National Tables, New York.
United Nations. 1975, 76, 78, 80, 1982. *Yearbook of Industrial Statistics*, New York.
United Nations Industrial Development Organization -UNIDO. 1996, 2003. International Yearbook of Industrial Statistics, Vienna. See http://www.unido.org
Valecillos H. 1990. *Estadísticas Socio-laborales de Venezuela*. Series Históricas, two volumes. BCV, Caracas.
Willebald H. 2011. *Natural Resources, Settler Economies and Economic Development during the First Globalization: Land frontier Expansion and Institutional Arrangements*. Ph.D. dissertation. Universidad Carlos III, Madrid.