THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École des Hautes Études en Sciences Sociales

Essais sur la Répartition du Revenu: Perspectives Méthodologiques,

Historiques et Institutionnelles avec Applications au Cas du Brésil (1926-2016)

Ecole doctorale n°465

ECOLE DOCTORALE ECONOMIE PANTHEON SORBONNE

Spécialité ANALYSE ET POLITIQUE ÉCONOMIQUES

COMPOSITION DU JURY :

Mme. LUSTIG Nora
Tulane University, Rapporteur

M. FERREIRA Francisco
The World Bank, Rapporteur

M. NOLAN Brian
University of Oxford, Membre du jury

M. MONNET Éric
Banque de France/PSE, Membre du jury

M. ALVAREDO Facundo
EHESS/Oxford/Conicet, Membre du jury

M. PIKETTY Thomas
EHESS/PSE, Membre du jury

Soutenue par Marc **MORGAN**
le 18 Décembre 2018

Dirigée par **Thomas PIKETTY**

L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES | PSL

ÉCOLE DES HAUTES ÉTUDES DE SCIENCES SOCIALES

PHD THESIS

to obtain the title of

Doctor of Philosophy of the École des Hautes Études en Sciences Sociales

Prepared and defended at the Paris School of Economics on 18 December 2018 by:

Marc MORGAN

Essays on Income Distribution

Methodological, Historical and Institutional Perspectives with Applications to the Case of Brazil (1926–2016)

Thesis Supervisor: Thomas PIKETTY

Jury

Nora LUSTIG (Tulane University – Referee)
Francisco FERREIRA (The World Bank – Referee)
Brian NOLAN (University of Oxford – Examiner)
Éric MONNET (Banque de France/PSE – Examiner)
Facundo ALVAREDO (EHESS/University of Oxford/Conicet – Examiner)
Thomas PIKETTY (EHESS/PSE – Supervisor)

Acknowledgements

This thesis is indebted to numerous individuals and organisations that have helped me in different ways throughout its realisation. I owe a sincere amount gratitude to the *Fundación Ramón Areces* for having financed the three academic years of my doctoral studies, without which I would not have been able to pursue the investigations here presented. I also thank the World Inequality Lab, and its funding partners – ERC (Grant 340831), Ford Foundation, INET (Grant INO14-00023), Alfrd P. Sloan Foundation (Grant WIL 1SUR 744), and other partners – for providing me with the necessary financial support over the last four months of my dissertation.

I gratefully acknowledge the institutions that accepted me as a masters student five years ago – the Paris School of Economics and the École des Hautes Études en Sciences Sociales. Both offered me excellent academic training and working conditions. I thank all the administrative and academic staff that aided me along the way.

I am deeply grateful to Thomas Piketty for his admirable supervision of my work and progress since my days as a masters student. I have benefited hugely from his transparency and sharp intuitions. His engagement with economics as a social science continues to be a constant source of inspiration.

Facundo Alvaredo equally guided me from the very beginning, being the person who initially orientated me to study Brazilian inequality. Without his honesty, his meticulous attention to detail and his knowledge of the Latin American context, my work would not have evolved for the better. Working with him was a great pleasure.

I was also fortunate to have crossed paths with Éric Monnet during my time as a masters student. I have learned a tremendous amount from the multiple exchanges we have had. His conception of economics and history has critically shaped my way of approaching research.

I equally thank Brian Nolan whose familiar spirit and passion for research on inequality and social policy also fed my work. I am grateful to have been invited by him to spend a fruitful semester at the Institute for New Economic Thinking (INET) at the University of Oxford earlier this year.

My referees, Nora Lustig and Francisco Ferreira, were easy choices. Their work on measuring social indicators and analysis of policy processes in Latin America and beyond have greatly influenced my studies. I have undoubtedly benefited from the conversations we exchanged at multiple conferences in the past.

I owe special thanks to Rodrigo Orair, who greatly facilitated my visit to the *Instituto de Pesquisa Econômica Aplicada* (IPEA) and the UN's IPC-IG in Brasília in 2017. My stay was immensely constructive for each chapter contained in this thesis, allowing me to directly experience the country I was studying, and to come into contact with exceptional researchers, particularly, Rodrigo himself, Sérgio Gobetti, Marcelo Medeiros, Pedro Souza, Fernando Gaiger Silveira, Sergei Soares, and others. I especially thank Pedro Souza, my co-author, with whom the crafting of the third chapter of this thesis was a thoroughly enjoyable and educational experience. I also received much encouragement from Eduardo Fagnani and his colleagues involved in the *Reforma Tributária Solidária* project during my visit to São Paulo in June, to which I am extremely thankful.

This thesis also owes a great deal to other fellow researchers and friends with whom I have interacted over the past years. At closest proximity, my colleagues in office R5-70 – Thomas Blanchet, my co-author, without whom the second chapter of this thesis would not have taken the form it has; Lydia Assouad; Clara Martinez-Toledano; Luis Estévez Bauluz; Aurélie Sotura; and Mathilde Muñoz; all of whom made daily office life a joy; without forgetting colleagues of my first office (Vanda Almeida, Alessandro Tondini and Giulio Iacobelli) – colleagues of the World Inequality Lab – Ignacio Flores, my co-author, from whom I benefited from innumerable discussions about various aspects of this thesis, and without whom the first chapter would not have seen the light of day; Mauricio de Rosa, a future co-author, after so many fruitful conversations on Latin American inequality and other topics; Amory Gethin, with whom I sharpened my perception of inequality and political cleavages in Brazil; Li Yang, Filip Novokmet, Léo Czajka, Yajna Govind, Daniel Waldenström, and Lucas Chancel, for their helpful feedback – and of PSE in general – Oscar Barrera and Marco Ranaldi, for their devoted interest and assistance in so many ways; as well as Santiago Garriga, Hector Moreno, Jaime Montana, Paolo Santini, José Montalban Castilla and many others. I particularly thank Gedeão Locks for engaging me further with Brazilian history and culture.

Finally, I thank my family – in particular, my parents, my brother, and my grandparents, for their influence, their support and their continuous interest in my pursuits. Lastly, to Naomi, with whom I share so much, for inspiring me every day, and for making this whole endeavour all the more enjoyable.

Summary

This PhD thesis consists of three essays on income distribution, from the point of view of statistical production (methods) and economic development (history and institutions), with applications to the case of Brazil. As Simon Kuznets alluded to in the 1950s, in order to begin to understand the relationship between inequality and development, we must first accurately measure the trends underlying the phenomena, and only then propose an analysis of the conditions giving rise to them. "With this as a beginning, we can then attempt to translate the elements of a properly understood past into the conditions of an adequately understood present".[1] This thesis attempts to purse this line of investigation.

Chapter 1, written with Thomas Blanchet and Ignacio Flores, presents the central methodological innovation of this thesis. It deals with the growing problem of household surveys to accurately portray the top tail if the income distribution. We propose a new flexible and rigorous method to reconcile survey data with information from more trusted sources, such as tax data. The method combines the two data sources while preserving the survey micro structure, ensuring the consistency of other socio-demographic variables at both the individual and aggregate levels. It thus allows future research to be carried out under a more representative distributive framework. The procedure is illustrated by empirical applications to five countries, covering both developed and less-developed contexts over numerous years.

Chapter 2 and Chapter 3 both make use of the method described in the first chapter to measure and analyse income inequality for different time periods and motives in Brazil – a perennial late-developing economy, where household surveys are an increasingly problematic source from which to obtain credible information on the relative income growth of different parts of the population.

Chapter 2 combines data from previously un-reconciled sources (surveys, tax data, social security records, national accounts) to uncover new evidence and a new understanding of income inequality in Brazil, focusing on the much debated period

[1] Kuznets, Simon. "Economic Growth and Income Inequality." *American Economic Review* 45, no. 1 (1955): 1-28.

of the 2000s. It finds that inequality within the Bottom 90% of the distribution declined, but concentration at the top persisted at very high levels. This dichotomy was given by the strong average income growth in both tails of the distribution mainly between 2002 and 2013, while the middle of the distribution was squeezed. The fall in inequality among the large mass of the population was due to the fall in labour income inequality, which was nonetheless insufficient to prevent the growing concentration of national income among economic elites. The chapter contextualises the findings to understand what may be driving the dynamics, from the progressive role of social policy, to the regressive role of the tax system and monetary policy.

Chapter 3, written with Pedro Souza, extends the Brazilian inequality analysis over a longer historical time-frame to examine where it has come from. The overall objective is to shed new light on long-run distributional dynamics and their connection with economic growth in a late-developing country. Based on the construction of a rich inequality dataset covering the whole population since 1976 and a top income group since 1926, and its combination with other distributional information and macro statistics, the chapter shows the unprecedented levels and persistence of income concentration in Brazil, despite tumultuous economic and political change. It explains the short-lived sustained egalitarian levelling in the country through an endogenous theory of institutional shifts, which originate from structural-economic changes, but get appropriated by elites to avoid the redistribution of fundamental factors (land, capital, income, education) that the economic changes and related social actors seem to demand. It identifies the military *coup* of 1964 and its aftermath as a crucial moment in the history of Brazilian inequality, whose ideas and policies largely suspended inclusive growth, consigned distributional bottlenecks to future governments, and whose legacy can still be perceived today.

Contents

Chapter 1

The Weight of the Rich: Improving Surveys Using Tax Data

Abstract: It is generally accepted that household surveys fail to accurately portray the top tail of the income distribution. Indeed, studies based on tax data challenge the credibility of surveys as a source to study inequality. To date, there is no broad consensus on how to best reconcile these two datasets. This chapter presents a novel method that enables a consistent combination of these two sources of data, under the assumption that tax data sets a credible lower bound on the amount of people with given levels of income. The resulting micro-dataset preserves the consistency of other socio-demographic variables at both the individual and aggregate levels. It thus allows researchers to analyze dimensions of social inequality under a more representative distributive framework. The procedure is illustrated by empirical applications to five countries, covering both developed and less-developed contexts over numerous years.

1 Introduction

For a long time, most of what we knew about the distribution of income came from surveys, in which randomly chosen households are asked to fill a questionnaire.[1] These surveys have been an invaluable tool for tracking the evolution of society. But in recent years, the research community has grown increasingly concerned with their limitations. In particular, surveys have struggled to keep track of income at the very top of the distribution.

For this reason, researchers have turned to a different source: tax data. The idea is not new; we can trace it back to the seminal work of Kuznets (1953), or even Pareto (1896). More recently, Piketty and Saez (2003) and Piketty (2003) applied their method to the latest data for France and the United States. This work was extended to more countries by many researchers whose contributions were collected in two volumes by Atkinson and Piketty (2007, 2010) and served as the basis for the World Inequality Database (`http://wid.world`).

But tax data has its own limitations. It covers only the top of the distribution, and includes at best a limited set of covariates. It is often not available as microdata but rather as tabulations summarising the distribution, which limits what can be done with them. The statistical unit that they use (individuals or households) depends on the local legislation and may not be comparable from one country to the next. This is why many indicators, such as poverty rates or gender gaps, still have to be calculated from surveys. The use of different — and sometimes contradictory — sources to calculate statistics on the distribution of income and wealth can make it hard to paint a consistent and accurate picture of inequality trends. This explains the ongoing effort to combine the different data sources at our disposal in a way that exploits their strengths, and corrects their weaknesses.

The Distributional National Accounts (DINA) project is a prominent example of this effort. Its guidelines (Alvaredo et al., 2017) emphasize the need to look at the entire distribution, harmonize concepts, and where possible to decompose by age and gender. Piketty et al. (2018) in the United States, and Garbinti et al. (2016) in France have used both survey and tax data to construct distributional statistics that account for all of the income recorded in national accounts. But these examples rely in large part on the existence of administrative microdata accessible to researchers, to which information from surveys can be added.

In many countries, both developed and less developed, such direct access is quite rare. Instead, we have tabulations of fiscal income, containing information on

[1] This chapter was written with Thomas Blanchet and Ignacio Flores.

the number and declared income of taxpayers by income bracket. The population coverage in the tabulations is often less than the total adult population, and the difference varies with the country studied. In such cases we have to proceed the other way round: rather than incorporating survey information into the tax data, we need to incorporate tax information into the survey data.

There has been a number of suggested approaches to deal with this problem, yet the literature has largely failed to converge towards a standard. In this paper, we develop a methodology that has significant advantages over previous ones, and which should cover most practical cases within a single, united framework. Our method avoids relying, to the extent possible, on *ad hoc* assumptions and parameters. We present a data-driven way to determine where the undercoverage of income starts in the survey data and beyond which point we merge incomes from tax data into the survey. We perform necessary adjustments in a way that minimize distortions from the original survey, and preserve desirable properties, such as the continuity of the density function. Rather than directly making assumptions on the behavior of complex statistics such as quantiles or bracket averages, our method makes easily interpretable assumptions at the level of observations. As a result, we can preserve the richness of information in surveys, both in terms of covariates and household structure. By looking at all variables simultaneously, we ensure the representativeness of the survey in terms of income while maintaining its representativeness in terms of age, gender, or any other dimension.

Our method proceeds in two steps, which are aimed at correcting for the two main types of error in surveys: non-sampling error and sampling error. Non-sampling error refers to issues that cannot easily be solved with a larger sample size, and typically arise from unobserved heterogeneous response rates. In the first step, we correct for these issues using a reweighting procedure rooted in survey calibration theory (Deville and Särndal, 1992). In doing so, we address a longstanding inconsistency between the empirical literature on top incomes in surveys, and the established practice of most survey producers. Indeed, since Deming and Stephan (1940) introduced their raking algorithm, statistical institutes have regularly reweighted their surveys to match known demographic totals from census data. Yet the literature on income has mostly relied on adjusting the value of observations, rather than their weight, to enforce consistency between tax and survey data. The theoretical foundations of this approach are less explicit and harder to justify.

This first step addresses non-sampling error, but it is limited in its ability to correct for sampling error, meaning a lack of precision due to limited sample size.[2] A

[2]Calibration methods can, to some extent, correct for sampling error. But their ability to do

radical example is the maximum income, which is almost always lower in the survey than in the tax data, something no amount of reweighting can do anything about. Top income shares of small income groups are also strongly downward biased in small samples (Taleb and Douady, 2015), so inequality will be underestimated even if all the non-sampling error has been corrected. To overcome this problem, we supplement the survey calibration with a second step, in which we replace observations at the top by a distribution generated from the tax data, and match the survey covariates to it. The algorithm for doing so preserves the distribution of covariates in the original survey, their correlation with income, and the household structure regardless of the statistical unit in the tax data. The result is a dataset where sampling variability in terms of income at the top has been mostly eliminated, and whose covariates have the same statistical properties as the reweighted survey. Because we preserve the nature of the original microdata, we can use the output to experiment with different statistical units, equivalence scales, calculate complex indicators, and perform decompositions along age, gender, or any other dimension included in the survey.

For practical use, we have developed a Stata command that applies the methodology described in this article. The program works with several types of inputs, ensuring flexibility for users. Our method may therefore easily be used by researchers interested in analysing the different dimensions of inequality.[3]

The remainder of the paper is structure as follows. In section 2 we relate our paper to the existing literature. In section 3 we lay out the theoretical framework of our method. This is followed by a practical guide of the method and its application to specific countries in section 4, before concluding.

2 Literature on Survey Correction Methods

Numerous studies have sought to combine administrative data and survey data primarily to improve the latter's representativeness or produce a more accurate distribution of income. We identify three distinguishable methodological strands present in this literature. The first strand opts to reweight survey observations. The second strand adjusts the income value of observations through a rescaling approach. Finally, a third strand identifies the need to employ a hybrid procedure by combining reweighting and rescaling.

so only holds asymptotically (Deville and Särndal, 1992), so it does not apply to narrow income groups at the top of the distribution.

[3]The packages to download are `bfmcorr` for the correction method, and `postbfm` for the postestimation output. Both commands come with a full set of user instructions.

2.1 Reweighting Observations

The papers that focus on reweighting survey observations generally tailor their approach to remedy the bias of nonresponse. Many studies in this literature rely on parametrically estimating a probabilistic model of response to adjust household survey weights without the use of external data sources on the distribution of incomes. Korinek et al. (2006) propose such an adjustment using the inverse of the probability of response for each household, which is estimated using nonresponse rates across geographic areas and the observable characteristics of respondents within regions. This type of approach, while not utilising auxiliary tax data, is sensitive to the degree of geographic aggregation used for inputting response rates into the adjustment. This is an issue explored in more detail by Hlasny and Verme (2017; 2018) for the U.S. and European case respectively, using similar probabilistic models. Depending on the nature of the survey data, greater or less geographic disaggregation on nonresponse rates can be more appropriate to the adjustment at hand. While the parametric models applied in these papers are data intensive, the estimations critically rely on observed survey distributions to adjust household weights given nonresponse rates across regions. Our proposal instead makes use of external administrative data, to guide us in how best to adjust household surveys, given the problem of nonresponse and its relation to income. This approach has the added benefit of indirectly tackling the problem of underreporting as we shall explain further on.

There are a few studies in this literature that combine surveys with external sources to measure inequality. An example of this is Alvaredo (2011), who for his second country-case study, on Argentina, estimates the corrected Gini coefficient by assuming that the top of the survey distribution (top 1% or top 0.1%) completely misses the richest individuals that are represented in tax data. This accounts for the bias of nonresponse and corrects the distribution via an implicit reweighting procedure. The specific form of the nonresponse bias that is assumed tacitly is, nonetheless, a rather restrictive one. Indeed, the correction implies a deterministic nonresponse rate equal to 1 above a previously selected fractile and 0 under it. Furthermore, in both of his empirical applications (on the U.S. and Argentina) the merging point is chosen arbitrarily.[4] Our method on the other hand tries at best to avoid arbitrary choices on the portion of the survey distribution to be corrected or on the form of the bias implied by the correction.

[4]In any case, the goal of the paper is not to tackle the nonresponse or misreporting biases directly, but to provide a simple estimation of a corrected Gini coefficient.

2.2 Rescaling Incomes

The general feature of this type of combination method is that it involves a rescaling of survey incomes with the tax incomes of equivalent rank. Although there is no unified theory or explicit justification behind most of these adjustment methods, they share some defining characteristics. In practice, they generally adjust distributions by replacing cell-means in the survey distribution of income with those from the tax distribution for the same sized cells (i.e. fractiles) in the population. The size of the cells varies by study (Burkhauser et al., 2016; Piketty et al., 2017; Chancel and Piketty, 2017; Czajka, 2017). Furthermore, the overall size of the population group whose income is to be adjusted is sometimes chosen arbitrarily, such as the top 20% (Piketty et al., 2017), top 10% in the distribution (Burkhauser et al., 2016; Chancel and Piketty, 2017), the top 1% (Burkhauser et al., 2016), or the top 0.5% of survey observations (DWP, 2015). It is also common to define the size of that group by choosing the point in the distribution beyond which the discrepancy between the average incomes in the two sources starts to become significant (Czajka, 2017). With a somehow different approach, Alvaredo (2011) uses tax data to adjust survey-based Gini coefficients, applying a method inspired from Atkinson and Piketty (2007) to the U.S. In constructing the corrected Gini, the top 1% in the income distribution from tax data directly replaces the top 1% from the survey. Thus survey incomes are rescaled accordingly.

Rescaling survey-respondents' declared income has been acknowledged as adjusting for the misreporting bias in surveys (Burkhauser et al., 2018; Jenkins, 2017). In section 5 of Appendix A we explain why this is only true under very strong and unrealistic assumptions, namely that the income rank in the survey distribution and in the true distribution are the same, and that underreporting is a deterministic function of this rank.

2.3 Combined Reweighting and Rescaling

Some voices stress the need to combine the aforementioned correction approaches. Bourguignon (2018), while reviewing the typical adjustment methods employed, correctly highlights that any method must dwell on three important parameters: the amount of income to be assigned to the top, the size of this top group, and the share of the population added to the top in the survey. The definition of these three parameters implies a correction procedure combining reweighting and rescaling. His analysis goes on to study the ways in which these choices impact the adjustments made to the original distribution data. However, this analysis does not shed light

on *how* to make these choices. Moreover, in reviewing multiple correction methods and applying them to Mexican survey data (including the combined case, where all three parameters mentioned take non-zero values), he only considers the situation "where nothing is known about the distribution of the missing income, unlike when tax records or tabulations are available" (Bourguignon, 2018). This is in contrast to our approach for correcting survey microdata, which combines the two previous methods, but which explicitly utilizes tax data, guiding users in how to best merge them with surveys to produce more realistic distributions of income.

To our knowledge the paper that comes closest to proposing an approach that resembles the one we propose here, in terms of criteria and methodology, is Medeiros et al. (2018) applied to Brazilian data. That is, it is the only study that combines tabulated tax data with survey micro-data by explicitly reweighting survey observations. More specifically, the authors apply a Pareto distribution to incomes from the tax tabulation to correct the top of the income distribution calculated from the census. Their method involves re-calibrating the census population by income intervals above a specified merging point, which is determined by the comparison of total income reported in the tax data and in the Census for the same intervals. The calibrating factors are based on the ratios between the populations in the same intervals of the two income distributions. However, while they increase the weight of observations above the merging point, they do not reduce the weight of individuals below this point, such that the corrected population ends up being larger than the original official population. This is an inconsistency our method avoids.

3 Theory and Methodology

To describe our method and the theory behind it, we will consider the following setting. Let X and Y be two real random variables. We will use Y to represent the true income distribution, part of which we assume is recorded in the tax data.[5] And we will use X to represent the income distribution recorded in the survey. Each random variable has a probability density function (PDF) f_Y and f_X, a cumulative probability function (CDF) F_Y and F_X, and a quantile function Q_Y and Q_X.

[5] In reality, part of the true income may also be missing from the tax data due to non-taxable income not reported in the declaration and tax evasion. But these issues are beyond the scope of this chapter.

3.1 Reweighting

In the first step, we adjust the weight of observations in the survey. In doing so, we are effectively adjusting the value of the survey density at different income levels. In this section we start by describing the intuition behind the correction in the simple univariate case. The next section explain how to use the theory of survey calibration to handle more complete settings.

3.1.1 Intuition

Let $\theta(y) = f_X(y)/f_Y(y)$ be the ratio of the survey density to the true density at the income level y. This represents the number of people within an infinitesimal bracket $[y, y + \mathrm{d}y]$ according to the the survey, relative to the actual number of people in the bracket. If $\theta(y) < 1$, then people with income y are underrepresented in the survey. Conversely, if $\theta(y) > 1$, then they are overrepresented.

The value of $\theta(y)$ may be interpreted as a relative probability. Indeed, let D be a binary random variable that denotes participation to the survey: if an observation is included in the sample, then $D = 1$, otherwise $D = 0$. Then Bayes' formula implies:

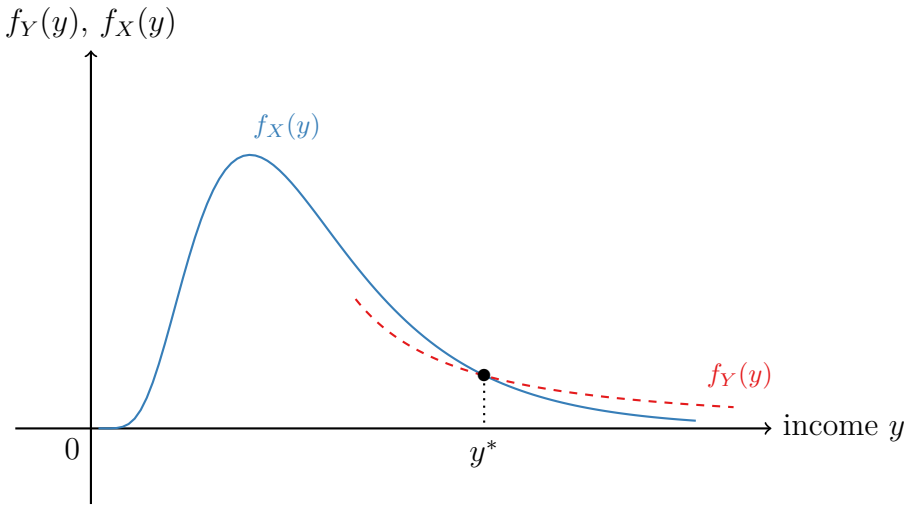
$$\theta(y) = \frac{f_X(y)}{f_Y(y)} = \frac{1}{f_Y(y)} \times f_Y(y) \frac{\mathbb{P}\{D = 1|Y = y\}}{\mathbb{P}\{D = 1\}} = \frac{\mathbb{P}\{D = 1|Y = y\}}{\mathbb{P}\{D = 1\}}$$

If everyone has the same probability of response, then $\mathbb{P}\{D = 1|Y = y\} = \mathbb{P}\{D = 1\}$, and $\theta(y) = 1$. Hence $f_X(y) = f_Y(y)$ and the survey is unbiased. What matters for the bias is probability of response at a given income level relative to the average response rate, which is why we have the constraint $\mathbb{E}[\theta(Y)] = 1$. Intuitively, if some people are underrepresented in the survey, then mechanically others have to be overrepresented, since the sum of weights must ultimately sum to the population size.

This basic constraint has important consequences for how we think about the adjustment of distributions. Any modification of one part of the distribution is bound to have repercussions on the rest. In particular, it makes little sense to assume that the survey is not representative of the rich, and at the same time that it is representative of the non-rich.

Figure 1.1 represents the situation graphically, in the more common case where $\theta(y)$ is lower for top incomes. We show a truncated version of f_Y since tax data often only cover a limited part of the whole distribution. The fact that the dashed red line $f_Y(y)$ is above the solid blue line $f_X(y)$ mean that top incomes are underrepresented. Therefore, lower incomes must be overrepresented, which is what we see below the point y^*. This pivotal value is unique assuming that θ is monotone. The appropriate

Figure 1.1: A "true" and biased income distribution



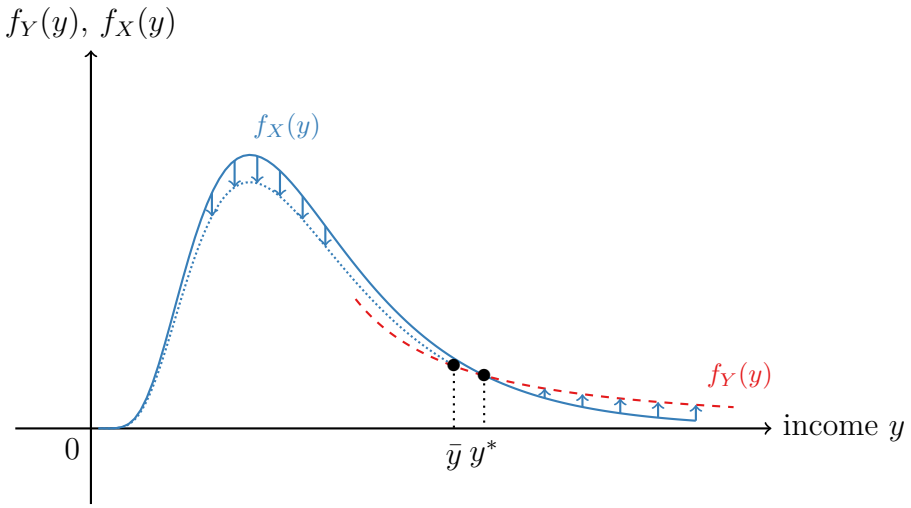
The solid blue line represents the survey density f_X. The dashed red line represents the tax data density f_Y, which is only observed at the top. For high incomes, the survey density is lower than the tax data density, which means that high incomes are underrepresented. If some individuals are underrepresented, then other have to be overrepresented: they correspond to people below the pivotal point y^*.

correction procedure here would be to increase the value of the density above it, and decrease its value below it. The intuition behind reweighting is that we have to multiply the survey density f_X by a factor $1/\theta(y)$ to make it equal to the true density f_Y. In practice, this means multiplying the weight of any observation Y_i by $1/\theta(Y_i)$.

When we observe both f_Y and f_X, we can directly estimate θ nonparametrically. But because we do not observe the true density over the entire support, we have to make an assumption on the shape of θ for values not covered by the tax data. We will assume a constant value. Behind this assumption, there are both theoretical motivations that we develop in section 3.2, and empirical evidence that we present in section 4. Intuitively, it means that there is no problem of representativeness within the bottom of the distribution, so that the overrepresentation of the non-rich is only the counterpart of the underrepresentation of the rich. We can therefore write the complete profile of θ as:

$$\theta(y) = \begin{cases} \bar{\theta} & \text{if } y < \bar{y} \\ f_X(y)/f_Y(y) & \text{if } y \geq \bar{y} \end{cases} \tag{1.1}$$

Figure 1.2: The intuition behind reweighting



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

We call $\bar{y}$ the *merging point*. It is the value at which we start to rely on the tax data. A naive choice would be to use the tax data as soon as they become available, but this will often lead to poor results. This is because the point from which the tax data become reliable is not necessarily sharp and well-defined, so in practice it will be better to start using the tax data only when it becomes clearly necessary. The proper choice of that point is an important aspect of the method on which we return to in section 3.1.2. For now we will take it as given, and only assume that it is below the pivotal point y^* of figure 1.1. Figure 1.2 shows how the reweighting using (1.2) operates.

Let $\tilde{f}_X$ be the reweighted survey, i.e. $\tilde{f}_X(y) = f_X(y)/\theta(y)$. By construction, we have $\tilde{f}_X(y) = f_Y(y)$ for $y \geq \bar{y}$. As indicated by upward arrows on the right of figure 1.2, the density has been increased for $y > y^*$. Since densities must integrate to one, values for $y < y^*$ have to be lowered. The uniform reweighting below $\bar{y}$ creates the dotted blue line.

3.1.2 Choice of the Merging Point

For many countries, tax data only covers the top of the distribution. We use the term *trustable span* to name the interval over which the tax data may be considered reliable. It takes the form $[y_{\text{trust}}, +\infty[$. This interval is determined by country specific

tax legislation: it is typically wider in developed countries than in less developed ones.

We do not usually wish to use the tax data over the entire trustable span. First, because the beginning of the trustable span is not always sharp. The reliability of the tax data increases with income in a way that is not well-defined, therefore it is more prudent to restrict their use to the minimum that is necessary. Second, once we are past the point where there is clear evidence of a bias, we prefer to avoid distorting the survey in unnecessary ways.

We call the *merging point* the value $\bar{y}$ at which we start using the tax data. We suggest a simple, data-driven way for choosing this point with desirable properties. In particular, we seek to approximately preserve the continuity of the underlying density function after reweighting. We start from the more simple case where $\bar{y}$ is inside the trustable span $[y_{\text{trust}}, +\infty[$, before moving on to consider cases where the trustable span may be too small.

Merging Point in the Trustable Span Assume that the bias function $\theta(y)$ follows the form (1.2):

$$\theta(y) = \begin{cases} \bar{\theta} & \text{if } y < \bar{y} \\ f_X(y)/f_Y(y) & \text{if } y \geq \bar{y} \end{cases} \tag{1.2}$$

We introduce a second function, the cumulative bias, defined as:

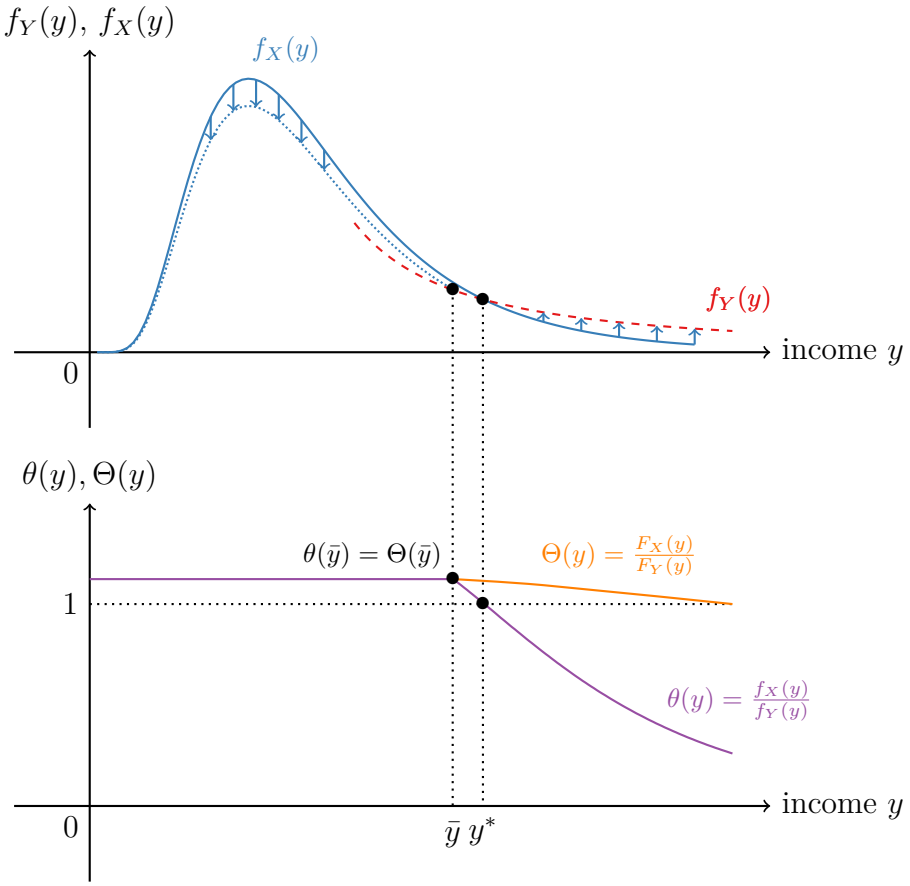
$$\Theta(y) = \frac{F_X(y)}{F_Y(y)} \tag{1.3}$$

In figure 1.3, we examine the shape of $\theta(y)$ and $\Theta(y)$ in relation to the density functions presented in figure 1.2.

We have the relationship $\Theta(y)F_Y(y) = \int_{-\infty}^{y} \theta(t) f_Y(t)\,\mathrm{d}t$. Given (1.2), for $y < \bar{y}$, $\Theta(y) = \bar{\theta}$. As figure 1.3 shows, we should expect the merging point $\bar{y}$ to be the highest value y such that $\Theta(y) = \theta(y)$.

We can contrast this choice of merging point with the one implicitly chosen in most rescaling approaches: the point at which the quantile functions of the survey and the tax data cross. This is equivalent to setting equal densities (i.e. $\theta(y) = 1$) until this merging point, which will in general be lower than ours. At the merging point, there is a discontinuity in $\theta(y)$ which jumps above one, and then progressively decreases toward zero. As a result, the people just above the merging point are implicitly assumed to be overrepresented compared to those below, even though they are richer. This discontinuity and lack of monotonicity of θ is hard to justify, and

Figure 1.3: Choice of Merging Point when $\bar{y} \geq y_{\text{trust}}$



our choice of merging point avoids it.

We can estimate both $\theta(y)$ and $\Theta(y)$ over the trustable span of the tax data. To determine the merging point in practice, we look for the moment when the empirical curves for $\Theta(y)$ and $\theta(y)$ cross, and discard the tax data below that point. This choice is the only one that can ensure that the profile of $\theta(y)$, and by extension the income density function, remains continuous.

The estimation of $\Theta(y)$ poses no difficulty as it suffices to replace the CDFs by their empirical counterpart in (1.3) to get the estimate $\hat{\Theta}_k$. For $\theta(y)$, however, we have to estimate densities. We define m bins using fractiles of the distribution (from 0% to 99%, then 99.1% to 99.9%, then 99.91% to 99.99% and 99.991% to 99.999%). We approximate the densities using histogram functions over these bins. This gives a first estimate for each bin that we call $(\tilde{\theta}_k)_{1 \leq k \leq m}$. The resulting estimate is fairly noisy, so we get a second, more stable one named $(\hat{\theta}_k)_{1 \leq k \leq m}$ using an antitonic (monotonically decreasing) regression (Brunk, 1955; Ayer et al., 1955; van Eeden,

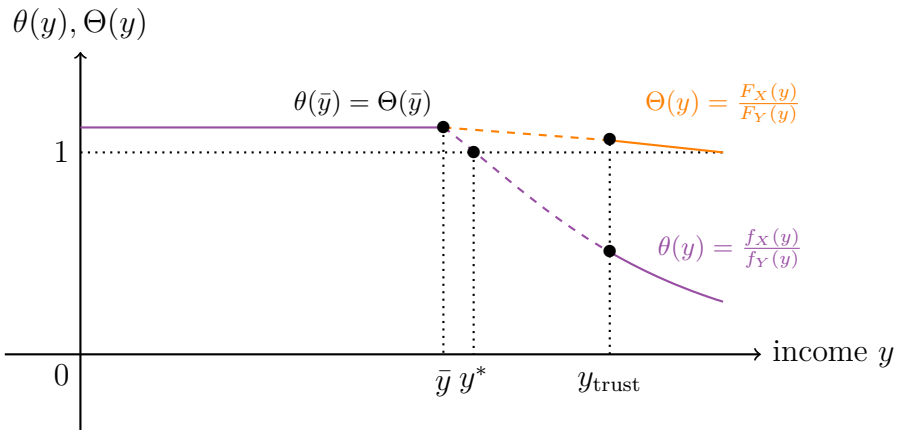
1958). That is, we solve:

$$\min_{\hat{\theta}_1,\dots,\hat{\theta}_m} \sum_{k=1}^{m} (\hat{\theta}_k - \tilde{\theta}_k)^2 \qquad \text{s.t.} \qquad \forall k \in \{2,\dots,m\} \quad \hat{\theta}_{k-1} \geq \hat{\theta}_k$$

We solve the problem above using the Pool Adjacent Violators Algorithm (Ayer et al., 1955). The main feature of this approach is that we force $(\hat{\theta}_k)_{1\leq k\leq m}$ to be decreasing. This turns out to be enough to smooth the estimate so that we can work with it, without the need introduce additional regularity requirements. We use as the merging point bracket the lowest value of k such that $\hat{\theta}_k < \hat{\Theta}_k$.

Merging Point Below the Trustable Span Sometimes the part of the distribution covered by the tax data is too limited to observe a merging point such that $\Theta(y) = \theta(y)$. This situation is represented in figure 1.4. Below y_{trust}, the value of $\theta(y)$ and $\Theta(y)$ have to be extrapolated until both curves cross, which is where we define the merging point.

Figure 1.4: Choice of Merging Point when $\bar{y} < y_{\text{trust}}$



We need to define a functional form for $\theta(y)$ in order perform the extrapolation (the value of $\Theta(y)$ follows from that of $\theta(y)$). We will assume the following:

$$\log \theta(y) = \gamma_0 - \gamma_1 \log y \tag{1.4}$$

which may also be written $\theta(y) = \mathrm{e}^{\gamma_0} y^{-\gamma_1}$. In addition to fitting the shape of the bias observed in practice, this form has the property of preserving Pareto distributions. Indeed, if $f_Y(y) \propto x^{-\alpha-1}$, then $f_X(y) = \theta(y) f_Y(y) \propto x^{-\gamma_1-\alpha-1}$, which is also a Pareto density. The parameter γ_1 may be interpreted as an elasticity of nonresponse: when the income of people increases by 1%, how much less likely are they to be represented in the survey.

While the equation (1.4) can be estimated by OLS, we need to take into account situations where tax data covers such a small share of the distribution that the number of data points is insufficient to estimate the regression reliably. Since the frontier between having and not having enough data is blurry, our preferred approach is to deal with the two cases at once using a ridge regression. The idea is that we can know from experience a typical value for γ_1 called γ_1^*. In the absence of data, it represents our baseline estimate.[6] As we observe new data, we may be willing to deviate from that value, but only to the extent that there is enough evidence for doing so. The ridge regression formalizes this problem as:

$$\min_{\gamma_0,\gamma_1} \sum_{i=1}^{m} (\log \tilde{\theta}_k - \gamma_0 - \gamma_1 \log y_k)^2 + \lambda(\gamma_1 - \gamma_1^*)^2$$

The first term is the same sum of squares as the one minimized by standard OLS. The second term is a Tikhonov regularization parameter that penalizes deviations from γ_1^*. If $m = 1$, then $\gamma_1 = \gamma_1^*$ and the sum of squares only determines the intercept. As we get more data points, the sum of squares gets more weight and results get closer to OLS. The parameter λ determines the strength of the penalization. The problem has an explicit solution expressible in matrix form (e.g. Hoerl and Kennard, 2000). We can have a Bayesian interpretation of the method where our prior for γ_1 is a normal distribution centered around γ_1^* and λ determines its variance. The solution of the ridge regression gives the mean value of the posterior. Once we have the estimation of γ_0, γ_1 we can simulate a tax data distribution by reweighting the survey data: the point at which $\theta(y)$ crosses $\Theta(y)$ becomes the merging point $\bar{y}$, and the reweighted survey from $\bar{y}$ to y_{trust} can be used to complete the tax data.

3.2 Calibration

3.2.1 General Setup

The previous section presented the main idea of the method. But while this intuition works well in the univariate case, the introduction of other dimensions from the survey (gender, age, etc.) complicates the problem significantly. Indeed, it is not enough for the survey to be solely representative in terms of income, we also need to preserve (or possibly enforce) representativeness in terms of these other variables.

Survey calibration is the standard tool to address these problems. The idea was introduced with the raking procedure of Deming and Stephan (1940). Deville and

[6] In practice, γ_1^* can be drawn from other "similar countries" that have sufficient data. For example, in our applications, we use the Brazilian γ_1^* to extrapolate the Chilean merging point (see section 4.2).

Särndal (1992) provided major improvements. While statistical institutes routinely use calibration methods with respect to age and gender variables, they are not yet traditionally used for income variables. We present the theory in its general setting below, before explaining how to apply it to the problem at hand.

Problem Survey calibration considers the following problem. We have a survey sample of size n. Each observation is a k-dimensional vector $\boldsymbol{x}_i = (x_{1i}, \dots, x_{ki})'$. The sample can be written $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, and the corresponding survey weights are $(d_1, \dots, d_n)$. We know from a higher-quality external source the true population totals of the variables $x_{1i}, \dots, x_{ki}$ as the vector $\boldsymbol{t}$. We seek a new set of weights, $(w_1, \dots, w_n)$, such that the totals in the survey match their true value, i.e. $\sum_{i=1}^n w_i \boldsymbol{x}_i = \boldsymbol{t}$.

This problem will in general have an infinity of solutions, therefore survey calibration introduces a regularization criterion to select the preferred solution out of all the different possibilities. The idea is to minimize distortions from the original survey data, so we consider:

$$\min_{w_1,\dots,w_n} \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i} \qquad \text{s.t.} \qquad \sum_{i=1}^{n} w_i \boldsymbol{x}_i = \boldsymbol{t} \tag{1.5}$$

That is, we minimize the χ^2 distance between the original and the calibrated weights, under the constraint on population totals: this is called linear calibration. While alternative distances are sometimes used, linear calibration is advantageous in terms of analytical and computational tractability.

Solution Solving the problem (1.5) leads to:

$$\frac{w_i}{d_i} = 1 + \boldsymbol{\beta} \boldsymbol{x}_i \tag{1.6}$$

where $\boldsymbol{\beta}$ is a vector of Lagrange multipliers determined from the constraints as:

$$\boldsymbol{\beta} = \boldsymbol{T}^{-1} \left(\boldsymbol{t} - \sum_{i=1}^{n} d_i \boldsymbol{x}_i \right) \qquad \text{with} \qquad \boldsymbol{T} = \sum_{i=1}^{n} d_i \boldsymbol{x}_i \boldsymbol{x}_i'$$

where the matrix $\boldsymbol{T}$ is invertible as long as there are no collinear variables in the $\boldsymbol{x}_i$ (meaning neither redundancy nor incompatibility of the constraints).[7] One undesirable feature of linear calibration is that it may lead to weights below one or even negative, which prevents their interpretation as an inverse probability and is incompatible with several statistical procedures. Therefore, in practice, we enforce

[7] In practice, we use the Moore–Penrose generalized inverse to circumvent the collinearity problem.

the constraints $w_i \geq 1$ for all i using an standard iterative method described in Singh and Mohl (1996, method 5). This is known as truncated linear calibration.

Interpretation There are two interpretations of the procedure. The first one is that of a nonresponse model. In this interpretation, the survey weights are the inverse of the probability of inclusion in the survey sample. The probability of inclusion is the product of two components. The first one depends on whether a unit is selected for the survey, regardless of whether that unit accepts to answer or not. We note $D_i = 1$ if unit i is selected, and $D_i = 0$ otherwise. The value $\delta_i = 1/\mathbb{P}\{D_i = 1\}$ is called the design weight. The design weight in constructed by the survey producer and therefore known exactly. The second component depends on whether a unit contacted for the survey accepts to answer or not. We note $R_i = 1$ if unit i accepts to participate in the survey, and $R_i = 0$ otherwise. The value $\rho_i = 1/\mathbb{P}\{R_i = 1\}$ is called the response weight. Since both D_i and R_i must be equal to 1 for a unit to be observed, the final weight is the product of these two components $\delta_i \rho_i$.

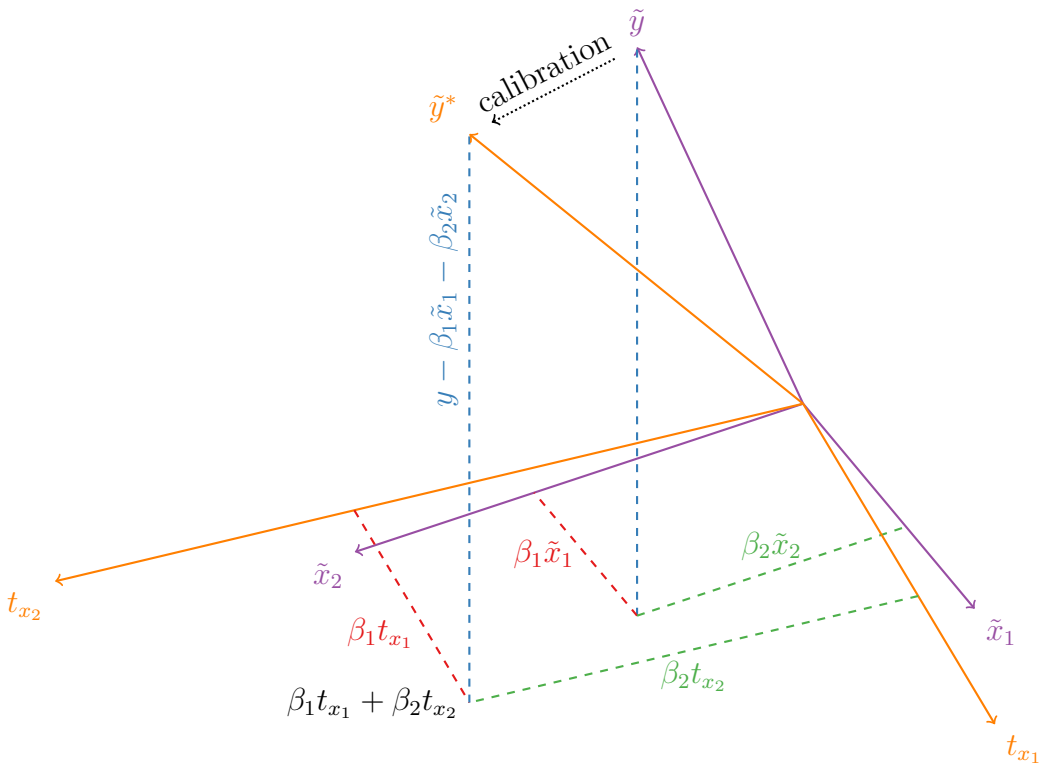
Nonresponse is unknown so it has to be estimated using certain assumptions. The simplest one is that ρ_i is the same for all units, therefore all weights are upscaled by the same factor so that their sum matches the population of interest. More complex models use information usually available to the survey producer, that is, basic sociodemographic variables which we will write $\boldsymbol{U}_i$. The survey producer models nonresponse as a function of these variables: $\rho_i = \phi(\boldsymbol{U}_i)$. The survey producer provides weights equal to $\delta_i \phi(\boldsymbol{U}_i)$. If nonresponse is also a function of income, which is not observed by the survey producer, then the estimated nonresponse will fail to accurately reflect true nonresponse, leading to biased estimates of the income distribution. Using the tax data $\boldsymbol{Y}_i$, we can estimate a new model that takes income into account: $\psi(\boldsymbol{U}_i, \boldsymbol{Y}_i)$. The final weight becomes:

$$\begin{aligned} w_i &= \frac{1}{\mathbb{P}\{D_i = 1\}} \frac{1}{\mathbb{P}\{R_i = 1\}} \\ &= \frac{1}{\mathbb{P}\{D_i = 1\}} \psi(\boldsymbol{U}_i, \boldsymbol{Y}_i) \\ &= \delta_i \phi(\boldsymbol{U}_i) \times \frac{\psi(\boldsymbol{U}_i, \boldsymbol{Y}_i)}{\phi(\boldsymbol{U}_i)} \\ &= d_i \times \frac{\psi(\boldsymbol{U}_i, \boldsymbol{Y}_i)}{\phi(\boldsymbol{U}_i)} \end{aligned} \tag{1.7}$$

Comparing equation (1.6) with (1.7), we see that the calibration problem suggests both a functional form and an estimation method for $\psi(\boldsymbol{U}_i, \boldsymbol{Y}_i)/\phi(\boldsymbol{U}_i)$. This functional form assumes nonresponse profiles that are as uniform (thus non-distortive) as

possible, and only modify the underlying distribution if it is necessary to do so. The preference for non-distortive functional forms can also help justify the use of a constant reweighting profile below the merging point in section 3.1.1.

Figure 1.5: Geometrical Interpretation of Linear Calibration



The survey totals $\tilde{y}$, $\tilde{x}_1$ and $\tilde{x}_2$ are shown in purple. The GREG estimator, which is equivalent to linear calibration, first projects $\tilde{y}$ onto $\tilde{x}_1$ and $\tilde{x}_2$ (dashed blue line). This projection is equal to $\beta_1\tilde{x}_1 + \beta_2\tilde{x}_2$. The true population totals t_{x_1} and t_{x_2} are in orange. We substitute them for $\tilde{x}_1$ and $\tilde{x}_2$ in the projection, which gives the value $\beta_1 t_{x_1} + \beta_2 t_{x_2}$. We add back the unexplained part of $\tilde{y}$ (dashed blue line) to get the calibrated total $\tilde{y}^*$.

The second interpretation is geometrical, and comes from the relationship between (1.5) and the generalized regression estimator (GREG). Assume that we seek to estimate the total of a survey variable y. We can directly use the survey total, which we will write $\tilde{y}$. But if we wish to exploit the information on the true population totals of the auxiliary variables $x_1, \ldots, x_k$, we can use the GREG estimator, whose logic is represented in figure 1.5. The idea is to first use the survey to project the variable of interest y onto the auxiliary variables $x_1, \ldots, x_k$ using an ordinary least squares regression. Hence we get a linear prediction $\hat{y}_i = \boldsymbol{\beta}\boldsymbol{x}_i$ of y_i, which corresponds to the part of y that can be explained by the auxiliary variables $x_1, \ldots, x_k$. We can then substitute the survey totals by their true population counterpart in the linear

prediction to get a new, corrected prediction of y. Adding back the unexplained part of y leads to the GREG estimator $\tilde{y}^* = \tilde{y} + \boldsymbol{\beta}(\boldsymbol{t} - \tilde{\boldsymbol{x}})$.

It can be shown algebraically that linear calibration is identical to the GREG procedure (Deville and Särndal, 1992). By using the calibrated weights, we systematically project the variable of interest on the calibration variables and perform the correction described above, without having to explicitly calculate the GREG estimator every time.

Application to Income Data The calibration problem is presented so as to enforce the aggregate value of variables. In order to use it to enforce the distribution of a variable, we have to discretize this distribution. In the case of income tax data, the income distribution may be presented in various tabulated forms, and we use the generalized Pareto interpolation method of Blanchet et al. (2017) to turn it into a continuous distribution.[8] We output the distribution discretized over a narrow grid made up of all percentiles from 0% to 99%, 99.1% to 99.9%, 99.91% to 99.99% and 99.991% to 99.999%. We discard tax brackets below the merging point whose choice is described in section 3.1.2. We then match the survey data to their corresponding tax bracket. It is in general necessary to regroup certain tax brackets to make sure that we have at least one (and preferably more) observations in each bracket. Otherwise the calibration will not be possible. We automatically regroup brackets to have a partition of the income distribution at the top such that each bracket has at least 5 survey observations.

Assume that we eventually get m brackets, with the k-th bracket covering a fraction p_k of the population. We create dummy variables $b_1, \dots, b_m$ for each income bracket. If the total population is N and the sample-size is n, then the calibrated weights should satisfy:

$$\forall k \in \{1, \dots, m\} \qquad \sum_{i=1}^{n} w_i b_{ik} = N p_k$$

Since these equations are expressed as totals of variables, they can directly enter the calibration problem (1.5). In practice, we are enforcing the income distribution through a histogram approximation of it.

The flexibility of the calibration procedure lets us put additional constraints in the calibration problem. In particular, if the survey is already assumed to be representative in terms of age or gender, then their distribution can be kept constant during the procedure. Hence we correct for the income distribution while main-

[8] See `wid.world/gpinter` for an online interface and a R package to apply the method.

taining the representativeness of the survey along the other dimensions. Additional constraints are also possible, if external information on other variables is available (see section 3.2.2).

For all the observations below the merging point, the dummy variables $b_1, \ldots, b_m$ are all equal to zero, so the weight adjustment only depends on a constant and possibly other calibration variables such as age and gender, but not income. This matches the uniform adjustment profile (1.2) at the bottom of the distribution that we present in section 3.1.1. The calibration, by construction, avoids distorting the bottom of the distribution because it is not necessary to enforce the constraints of the calibration problem.

Our correction procedure also constrains the number of times the weights are expanded or reduced to avoid disproportionate adjustments to single observations already in the dataset. Consequently we introduce the condition that brackets with a $\theta(y)$ outside the boundary defined by $1/n \leq \theta(y) \leq n$ are automatically grouped into larger brackets. The default limit we choose (which can be changed by users) is n = 5. Thus, in this case, no observation would have their weight multiplied by more than 5 times or less than 0.2 times.

3.2.2 Extensions

The calibration framework is generic enough to incorporate information into the survey in different forms. While the most standard problem is to directly correct the income distribution using the income concept of interest, more complicated settings can sometimes occur. The flexibility of the calibration framework makes it generally possible to deal with these settings without resorting to additional *ad hoc* assumptions. We discuss below three common cases.

Using Population Characteristics by Income Tax data can provide information on the population characteristics by income level, typically, the gender composition. This can tell us how the interaction between income and other characteristics impacts the bias, so it can be useful to include this information in the survey.

Assume that we have m income tax brackets that contain a share $p_1, \ldots, p_m$ of the overall population N. For each of them, we know the share $\boldsymbol{s} = (s_1, \ldots, s_m)$ of people with a given characteristic, such as belonging to a certain gender or age group. Let d_i be the variable equal to 1 if unit i belongs to that group in the survey, and 0 otherwise. Let b_{ik} be the variable equal to 1 if unit i in the survey is in income bracket k, and 0 otherwise.

To make sure that the survey reproduces the information in the tax data, we add the following constraints to the calibration problem (1.5):

$$\forall k \in \{1, \dots, m\} \qquad \sum_{i=1}^{n} w_i b_{ik} d_i = N s_k p_k$$

Using Income Composition Another source of information that is commonly available in tax data is the composition of income within brackets. Using that information is useful if we assume that the bias may be different for people that derive their income from, say, capital rather than labor.

Assume that we have m income brackets. For each of them, we know the share $\boldsymbol{s} = (s_1, \dots, s_m)$ of capital income. In the survey, total income is recorded as y_i and capital income as c_i. Let b_{ik} be a variable equal to 1 if unit i in the survey is in income bracket k. In order to enforce the constraint that the share of capital income within each bracket is the same as in the tax data, it suffice to enforce the constraints:

$$\forall k \in \{1, \dots, m\} \qquad \sum_{i=1}^{n} w_i b_{ik} (c_i - s_k y_i) = 0$$

Indeed, the first part of the sum is $\sum_{i=1}^{n} w_i b_{ik} c_i$, which is the total capital income of the bracket. In the second part we have the total income of the bracket $\sum_{i=1}^{n} w_i b_{ik} y_i$, multiplied by the capital share s_k. This constraint can be expressed as a total of the variable $b_{ik}(c_i - s_k y_i)$. We can see that units will see their weight decrease or increase depending on whether their capital share is below or above the average of the bracket they belong to.

Using several income concepts Until now we have considered the case where the income recorded in tax data more or less matches the income concept of interest, and the income likely to drive the bias. Yet sometimes only part of the income is recorded in the tax data. For example, in developing countries, only income from the formal sector may be recorded in the tax data, and there is a sizeable informal sector only present in the survey data, which is widely spread across the distribution (as in Czajka (2017)).

In such cases, it would be problematic to directly apply the calibration method described previously. Indeed, since the adjustment factor of the weights would only depend on formal sector income, two people with the same income, one working in the formal sector and the other in the informal sector, would see their weight adjusted very differently. As a result, there would be almost no correction for the income distribution of the informal sector.

The solution to that problem is to use Deville's (2000) generalized calibration

approach. The standard calibration approach formulated in (1.5) does not specify on what variable the weight adjustment factors should depend. In the solution of the problem, they depend directly on the variables used in the constraint. This is because the method always favors the least distortive adjustments, so it only uses the variables most directly related to the constraints.

If we have some prior knowledge of what the bias should depend on, then we can use generalized calibration to specify these variables *ex ante*. We still use $\boldsymbol{x}_i$ to denote the k calibration variables for which we know the true population totals $\boldsymbol{t}$. In the example, it would include formal sector income in addition to basic sociodemographic characteristics. We also define $\boldsymbol{z}_i$, a vector of instrumental calibration variables with the same size as $\boldsymbol{x}_i$. They may include variables in $\boldsymbol{x}_i$ (e.g. sociodemographic variables) but more importantly also some variables imperfectly correlated with the $\boldsymbol{x}_i$, in the example the sum of formal and informal sector income. We write the calibration problem as finding $w_1, \dots, w_n$ such that:

$$\sum_{i=1}^{n} w_i \boldsymbol{x}_i = \boldsymbol{t} \qquad \text{and} \qquad \forall i \in \{1, \dots, n\} \qquad \frac{w_i}{d_i} = 1 + \boldsymbol{\beta} \boldsymbol{z}_i \tag{1.8}$$

When $\boldsymbol{x}_i = \boldsymbol{z}_i$, the problem (1.8) is equivalent to (1.5). The solution of (1.8) given by Deville (2000) is similar to that of (1.6):

$$\boldsymbol{\beta} = \boldsymbol{T}^{-1} \left(\boldsymbol{t} - \sum_{i=1}^{n} d_i \boldsymbol{x}_i \right) \qquad \text{with} \qquad \boldsymbol{T} = \sum_{i=1}^{n} d_i \boldsymbol{z}_i \boldsymbol{x}_i'$$

We can understand the name "instrument" for the $\boldsymbol{z}_i$ by going back the the GREG estimator (see figure 1.5 and section 3.2). While we may view the standard calibration as performing a projection of the variable of interest y_i onto the calibration variables $\boldsymbol{x}_i$ using an OLS regression, the generalized calibration performs that same projection using an IV regression with $\boldsymbol{z}_i$ as a vector of instruments for $\boldsymbol{x}_i$. For this to work properly, we need $\boldsymbol{z}_i$ to be sufficiently correlated with $\boldsymbol{x}_i$, otherwise we face a weak instrument problem similar to that of traditional IV regressions (Lesage et al., 2018). This is not a major concern in the example since the sum of formal and informal income is strongly correlated with formal income by construction.

3.3 Replacing and Matching

After applying the methods of section 3.1, the survey should be statistically indistinguishable from the tax data. However, the precision that we get at the top of the income distribution may still be insufficient for some purposes. Indeed, the number of observations in the survey is still significantly lower than what we would

get in theory from administrative microdata. The extent to which this represents a problem varies. If we use survey weights to, say, run regressions and get correct estimates of average partial effects in presence of unmodeled heterogeneity of effects (Solon et al., 2015), then the reweighting step is enough. But problems may arise if we wish to produce indicators of inequality, especially the ones that focus on the top of the distribution, like top income shares. The combination of a low number of observations with fat-tailed distributions can create small sample biases for the quantiles and top shares (Okolewski and Rychlik, 2001; Taleb and Douady, 2015), and skewed distributions of the sample mean (Fleming, 2007). In most cases, we would underestimate levels of inequality.

Unlike problems caused by, say, heterogeneous response rates, these biases are part of *sampling error*. They do not reflect fundamental issues with the validity of the survey, but arise purely out of its limited sample size. The calibration method (section 3.2) does, to some extent, reduce sampling error. Yet it only does so under asymptotic conditions (Deville and Särndal, 1992) that cannot hold for narrow groups at the top of the income distribution. For this reason, we prefer to consider that the role of survey calibration in our methodology is to deal with *non-sampling error*. We use a different approach to deal with sampling error.

In particular, we aim to solve the case where tax statistics include a positive number of income-declarations beyond the survey's support. That is, we need to account for individuals declaring higher income than the richest persons in the surveys, which cannot be solved by re-weighting observations. To do so, we start from the original tax tabulations, which were created from the entire population of taxpayers and should therefore be free of sampling error. We use it to estimate a continuous income distribution (Blanchet et al., 2017) that reproduces the features of the tax data with high precision. We then match statistically the information in the calibrated survey data with the tax data by preserving the rank of each observation.

First, we inflate the number of data points in the survey by making k_i duplicates of each observation i. We attribute to each new observation the weight $q_i = w_i/k_i$, where w_i is the calibrated weight from the previous step. We choose $k_i = [\pi \times w_i]$ where $[x]$ is x rounded to the nearest integer. Therefore all new observations have an approximately equal weight close to $1/\pi$. The size of the new dataset, made out of the duplicated observations, can be made arbitrarily high by adjusting π, yet any linear weighted statistic will be the same over both datasets.

Let M be the number of observations in the new dataset. The weights are assumed to sum to the population size N. We will associate to each of them a small share $[0, q_{j_1}/N], [q_{j_1}/N, (q_{j_1} + q_{j_2})/N], \ldots, [\sum_{k=1}^{M} q_{j_k}/N, 1]$ of the true population. If

we attribute to each observation the average income of their population share in the tax data, then by construction the income distribution of the newly created survey will be the same as in the tax data. We rank observations in increasing order by income to preserve the joint distribution (i.e. empirical "copulas") between income and the covariates in the survey.

From an intuitive perspective, this process can be described as replacing the income of observations beyond the merging point with the income of observations with equivalent weight and rank in the tax distribution. This step ensures that the we reproduce exactly the income distribution from tax data, preserve the surveys' covariate distribution (including the household structure), and limit distortions in the relationship between income and covariates from survey data.

4 Applications

Our method can be replicated for all countries with the requisite data, namely, survey micro-data covering the entire population and tax data covering at least a fraction of it.[9] In order to illustrate how the method operates in practice, we apply it to data from five countries, three developed (France, U.K., Norway) and two less-developed (Brazil, Chile). Our chosen case studies showcase the wide applicability of the method to both developed countries and less developed ones, whose data quality is more challenging.

4.1 Definitions and Data

A crucial preliminary step in the analysis is to reconcile both the definition of income and the unit of observation in national surveys with the ones that are used in tax declarations. Our algorithm functions under the supposition that these definitions have been made consistent in the two datasets.[10] For European countries our analysis broadly covers the years 2004-2014. For Brazil, we cover 2007-2015 and for Chile we include the years 2009, 2011, 2013 and 2015.

[9] In the case where users only avail of tabulated survey data our method will still perform the correction, using percentile bracket-information from the synthetic micro-files produced by the *gpinter* program (see `wid.world/gpinter`).

[10] The main purpose of our method is to ensure the representativeness of top incomes in surveys using tax data. Nonetheless, the procedure also preserves the representativeness of other variables for which the survey is assumed to be already representative. These typically concern gender and age variables. Our calibration process leaves the distribution of these variables, or any other specified categorical variables, unchanged.

Income Concept Given that we seek to approximate the benchmark distribution, our method is by definition anchored to the income concept that is used in the tax tabulations, which in all of our case studies is pre-tax income. However, countries differ in the income concept included in their respective surveys. Brazil's PNAD reports individuals' pre-tax income, while Chile's CASEN gives after-tax income. The latter situation thus requires an imputation of taxes paid to arrive at gross incomes. Appendix A explains how this imputation is done for the Chilean case, as well as the construction of income units in surveys and their approximation with tax data in all countries. For the European countries we work with gross incomes (pre-tax and employee contributions deducted at source) from the SILC database. France is the exception since incomes reported in the tax files are net of social contributions deducted at source. For this reason we use the concept of net income in SILC for France that deducts social contributions levied at source.

The tax data we use is presented in tabulated form, containing at the very least, the number of income recipients by given income intervals and the total or average income declared within each interval. For France, we use the tabulated tax statistics produced by Garbinti et al. (2016) from the ministry of finance's tax microdata. The data cover all tax units (*foyers fiscaux*, singles or married couples), with about 50% of these subject to positive income tax. For the U.K. we use tax tabulations from the Survey of Personal Incomes (SPI) available from the Office of National Statistics. The underlying data covers about 80-90% of tax units (individuals) aged 15+, with about 60% subject to positive income tax. For Norway, we use tax data from Statistics Norway, which covers 100% of tax units (individuals) aged 17 and over, of which roughly 90% have postive income tax payments. For Brazil we use tax data from the personal income tax declarations (DIPRF tables), which covers about 20% of the adult population, with about 14% subject to the personal income tax on taxable income. For Chile we exploit income tax data from the *Global Complementario* and *Impesto Único de Segunda Categoría* (IGC and IUSC tabulations), which covers 70% of the adult population, with about 20% subject to the personal income tax on taxable income.

Observational Unit Concerning the observational units, we anchor the definition to the official tax unit in each country. In all of our country cases declarations are made at the individual level, except in France and Brazil, where declarations are jointly filed by married couples (in the case of the latter, at their own discretion). However, for France we make use of the individually-declared fiscal income files produced by Garbinti et al. (2016). Therefore for all countries, we define the unit of

analysis across datasets as individual income, including for Brazil, where the joint income of couples is equally split between the component members (see Appendix A and chapter 2 for further details).

4.2 Empirical Bias and Corrected Population

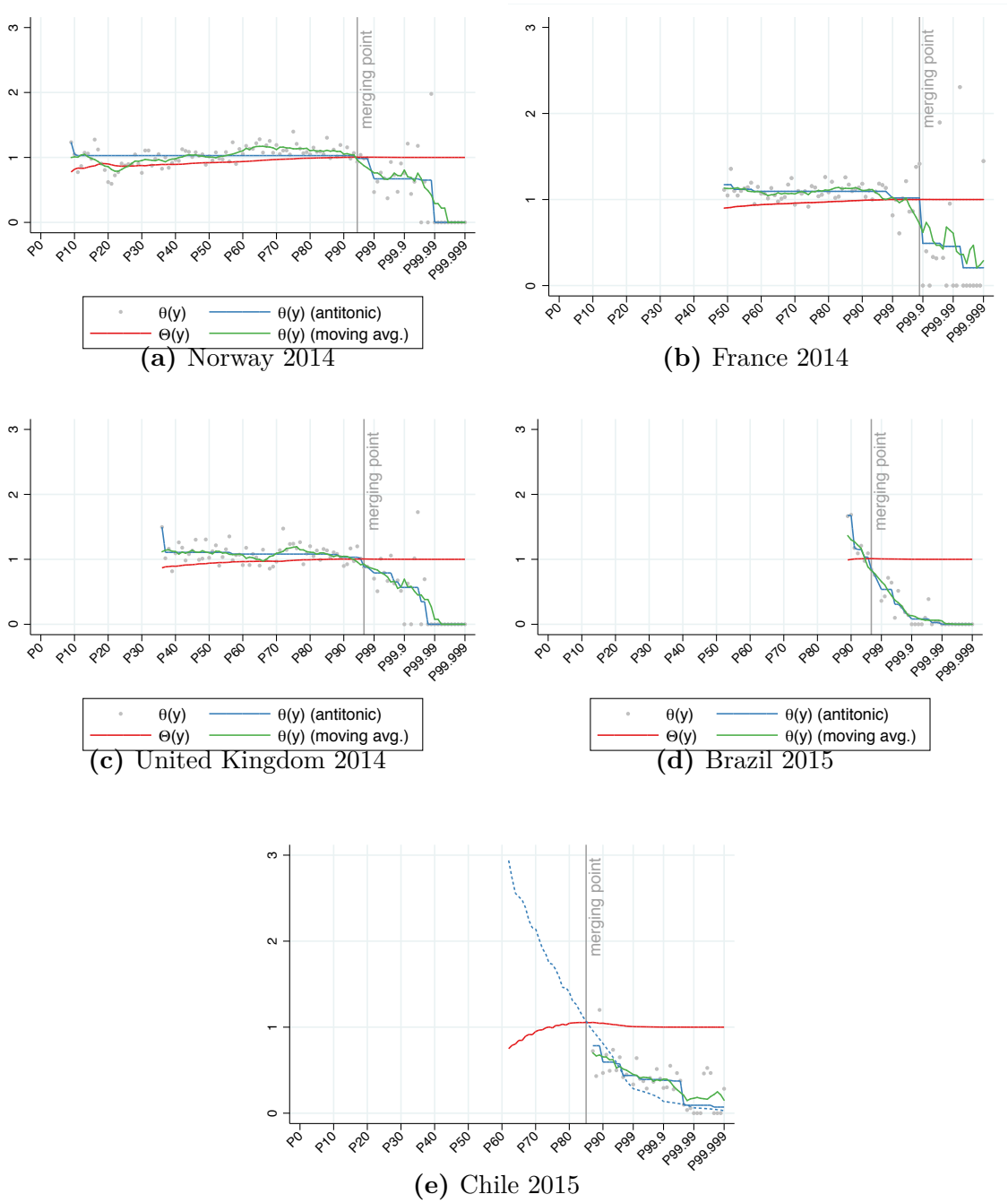
The Shape of the Bias Our method proposes to find the merging point between surveys and tax data by comparing the population densities at specified income levels, as explained in section 3.1.2. To do so we first interpolate the fiscal incomes in the tabulation using the generalized Pareto interpolation (`https://wid.world/gpinter`) developed by Blanchet et al. (2017), which allows for the expansion of the tabulated income values into 127 intervals.[11] Using the tresholds of these intervals we can construct our key statistics $\theta(y)$ and $\Theta(y)$.

Figure 1.6 presents depictions of the shape of the empirical bias within the tax data's "trustable span" for all countries for the latest available year. First of all, the shape of the bias we measure from the data is very similar to the one we present in the theoretical formalization, depicted in Figures 1.3 and 1.4. In particular, we always observe a convex shape in the top tail, to the right of the merging point. It thus appears that surveys tend to increasingly underestimate the frequency of incomes beyond a certain point in the distribution. For the more developed countries (Norway, France and the United Kingdom), the shape of the empirical bias $\theta(y)$ can be observed for a more comprehensive share of the population, due to their greater population coverage in tax data. This enables us to empirically test our theoretical expectations on the specific behavior of the bias to the left of the merging point. We indeed observe on the left side of Figures 1.6a 1.6b 1.6c, a general stability in the relative rate of response, with averages trending above 1. The extent and quality of tax data below the merging point in less developed countries is such that we cannot observe the same trends.[12] The merging points found by our algorithm vary by country and by year, again revealing differences in data quality and coverage between them. The Chilean case (Figure 1.6e) provides an example of our program

[11] These comprise of 100 percentiles from P0 to P100, where the top percentile (P99–100) is split into 10 deciles (P99.0, P99.1, ..., P99.9-100), the top decile of the top percentile (P99.9–100) being split into ten deciles itself (P99.90, P99.91, ..., P99.99-100), and so forth until P99.999. This interpolation technique, contrary to the standard Pareto interpolation, allows us to recover the income distribution without the need for parametric approximations. It estimates a full set of Pareto coefficients by using a given number of empirical thresholds provided by tabulated data. As such the Pareto distribution is given a flexible form, which overcomes the constancy condition of standard power laws, and produces smoother and more precise estimates of the distribution.

[12] Tax enforcement issues affecting this portion of the distribution could be at play here, as well as the sharp difference in incomes between the top and the rest in these countries leading to higher inequality levels than developed countries.

Figure 1.6: Merging Point in 6 Countries, Latest year



(a) Norway 2014

(b) France 2014

(c) United Kingdom 2014

(d) Brazil 2015

(e) Chile 2015

Notes: the figures depict the estimated bias in the survey relative to the tax data. Grey dots are, for each quantile of the fiscal income distribution, the ratio of income density in the survey over that of tax data. The green line is the centered average of $\theta(y)$ at each quantile and eight neighboring estimates. The blue line is the result of an *antitonic* regression applied to $\theta(y)$. It is constrained to be decreasing as it is used to find a single merging point. The blue dotted line, which only appears in figure 1.6e,is an extrapolation of the trend described by $\theta(y)$ based on a *ridge* regression. The red line is the ratio of the cumulative densities. For details refer to section 3.1.2.

needing to extrapolate the shape of the bias to find the merging point (see Section 3.1.2) For this case we rely on parameters observed for Brazil (specifically, values for the elasticity of response to income) above its trustable span as inputs for the Chilean extrapolation.[13] The fit with the existing data seems to work quite well. The empirical bias that is observed in previous years for all countries is presented in section 2 of Appendix A

Table 1.1: Structure of Corrected Population: Latest Year

Country	Population over Merging Point (% total population)		Corrected population		
	Tax data	Survey	Total	Share inside survey support	Share outside survey support
	[2]	[3]	[4] = [2] − [3]	[5]	[6]
Chile	14.0%	9.2%	4.8%	99.99%	0.01%
Brazil	3.0%	1.9%	1.1%	98.2%	1.8%
UK	3.0%	2.5%	0.5%	93.6%	6.4%
Norway	5.0%	4.6%	0.4%	96.0%	4.0%
France	0.1%	0.05%	0.05%	99.0%	1.0%

Notes: The table orders countries by the size of the corrected population. Column [2] shows the proportion of the population that is above this merging point in the tax data. Column [3] shows the proportion that is above the merging point in survey data. The difference between the two is the proportion of the survey population that is corrected (Column [4]). As explained in the text, we adjust survey weights below the merging point by the same proportion. The corrected proportion above the merging point can be decomposed into the share of the corrected population that is inside the survey support (up to the survey's maximum income) and the share that is outside the support (observations with income above the survey's maximum). Brazil and Chile refer to 2015, while all the European countries refer to 2014.

Corrected Population Our program then adjusts the individual weights of survey respondents in line with information from tax data, as described in section 2.1. We provide some summary statistics of the population we correct in Table 1.1, again using the last available year for each country as illustrations (see section 3 of Appendix A for other years). According to the comparison of surveys with tax records, a varying proportion of the total population is adjusted at the top of the survey distribution in each country (column [4] of Table 1.1), ranging from 4.8% in Chile to 0.05% in France for their most recent years.[14] This is derived from the comparison of the share of the population above the merging point in the two datasets. Since we use

[13] The value of the baseline elasticity of response to income, γ_1^*, extracted from the Brazilian data is -0.99.

[14] Across years there is less variation in this share, with Norway and particularly France being relative exceptions. In the French case, we believe the significant break in the series is due to the use of register data in SILC alongside the household survey from 2008. Despite the SILC survey making use of register data for countries like France and Norway, the goal is not to over-sample the top of the distribution, but rather to improve the precision of responses.

incomes in tax data as the benchmark for the top of the distribution, the share of the population above the merging point in tax data is directly related to the merging point. The share of the population above this point in surveys is always lower, indicating under-coverage of top incomes. But in both cases, the overwhelming majority of the adjustment (over 90%) can be seen to come from inside the survey support, rather than outside the survey's original support. In general, this step of the algorithm should be a useful guide for researchers to assess the income coverage of surveys across countries. For instance, it appears on the basis of our analysis that the Brazilian surveys do a better job at capturing gross income, given the lower share of the underrepresented population, than the Chilean household surveys. Moreover, comparing France and the UK, it seems that sampling error is greater in the UK surveys, given the higher share of the population beyond the survey's maximum income that needs to be added. Non-sampling error itself is greatest in Chile and France, derived from the share of the corrected population found inside the survey's support.

4.3 Income Distribution

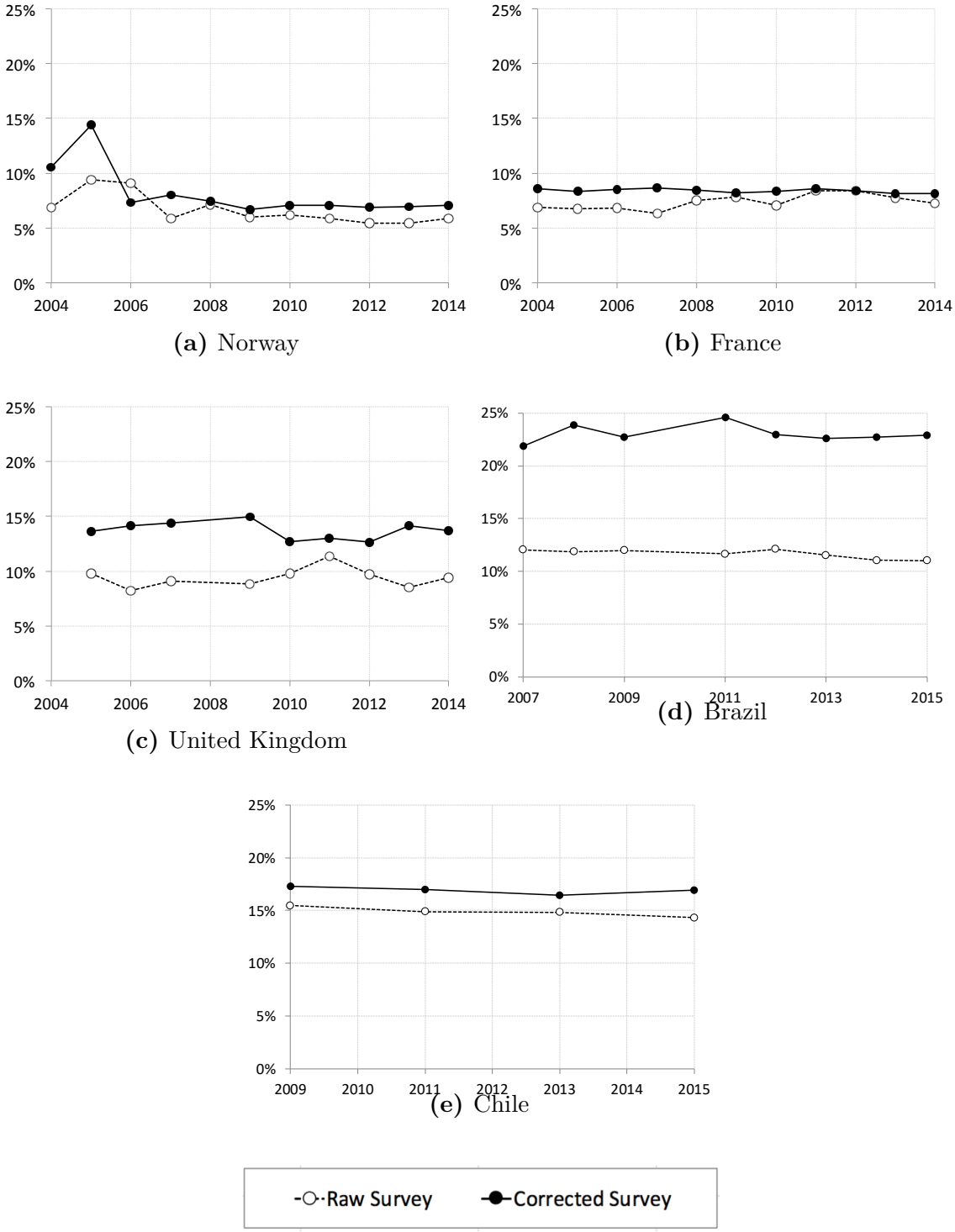
As detailed above, our method produces an adjusted micro dataset that maintains the survey's original design along a more representative income distribution. We can unveil how this merged distribution changes with respect to the raw survey distribution.

Top Income Shares Our adjustment procedure generally makes significant upward corrections to the shares of income going to the top of the distribution in the surveys. The size of the adjustment, however, varies with countries. Figure 1.7 depicts this for the Top 1% share in 5 countries for all years with data available.[15] Brazil has the most extensive one, with a top 1% share that increases about 10 percentage points every year (Figure 1.7d). Conversely, France and Norway experience relatively smaller adjustments, starting from relatively lower levels of inequality.

The quality of both surveys and tax statistics may have a substantial impact in the size of the adjustment. For instance, in the case of France, several improvements were made to the survey's methodology since 2008. In particular, the matching of individuals across survey and tax statistics allowed the use of tax data as an external source to assess individual income without recourse to self-reporting. Visibly,

[15]The one exception to this upward correction is Norway in 2006 (see Figure 1.7a). However, this is likely due to a change in the local tax legislation affecting the distribution of business profits (Alstadsæter et al., 2016).

Figure 1.7: The Top 1% Share Before and After Correction



(a) Norway

(b) France

(c) United Kingdom

(d) Brazil

(e) Chile

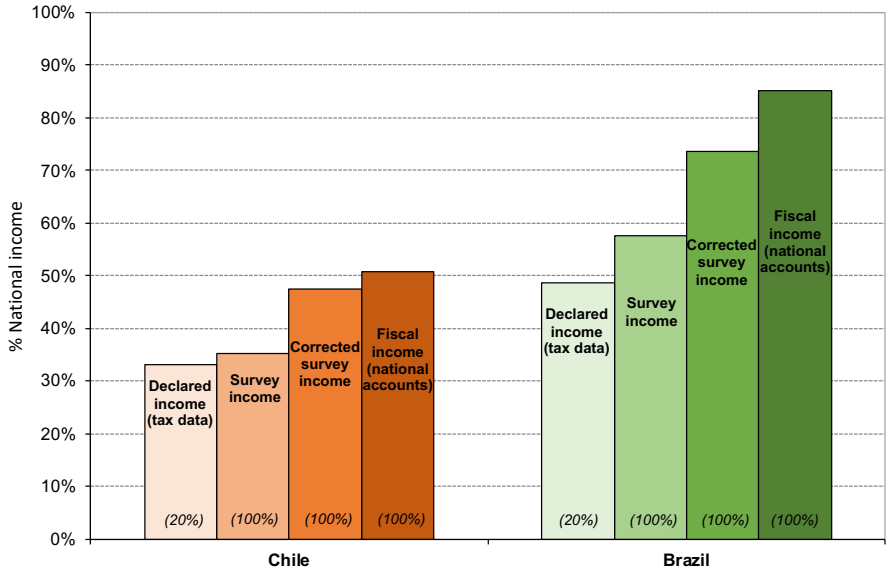
in Figure 1.7b the gap between raw and corrected estimates is reduced from 2008 because the size of the survey bias was reduced with the methodological novelties.

Moreover, when we compare the size of the adjustment in Chile and Brazil (Figures 1.7d and 1.7e respectively), two highly unequal Latin-American countries, the latter has a considerably higher adjustment. One of the reasons that could be behind this phenomenon is the fact that capital income, especially dividends, is better recorded in Brazilian tax statistics. Indeed, the Brazilian tax agency has relatively good means to verify the accuracy of capital income declarations (see chapter 2), while Chilean tax authorities are generally constrained by bank secrecy (Fairfield and Jorratt De Luis, 2016). In this case, the limited quality of Chilean tax statistics explains the smaller correction [16] Following the same rationale, the inclusion or exclusion of some types of income in a given dataset can also affect the size of the correction. In the case of Norway, tax incentives started favoring the retention of corporate profits inside corporations after 2005, with the creation of a permanent dividends tax in 2006. This resulted in less dividend payments, and thus less income to be registered as personal income in tax data. The reform also gave strong incentives for higher-than-normal dividend payouts in 2005, which contributed to the sharp increase in top shares observed for this year (Aaberge and Atkinson, 2010; Alstadsæter et al., 2016). In Figure 1.7a, it can be clearly perceived that the size of the adjustment appears to drop durably after this year. Additionally, it should be noticed that the Norwegian survey appears to be rather insensitive to this change, implying that dividends where badly represented before 2005.

Another potential explanation for the difference in the size of adjustments could be the difference in levels of inequality between countries. This could help explain for instance, why the survey in the United Kingdom receives an adjustment that is higher than the one of both Norway and France, but lower than the one of Brazil (Figure 1.7c). In addition, Brazil offers the clearest illustration of the distinct trends in inequality that can emerge after making a correction to the survey's income representation. While the raw survey depicts falling top income shares, the corrected survey distribution returns stable if not slightly increasing top shares. Distinct trends are also visible, albeit for shorter periods of time, in the other countries.

[16] There is also a considerable difference between these countries' tax systems and their respective incentives. In Chile most dividends received by individuals are taxed, while in Brazil they are not. This, in addition to the fact that Chilean realized capital gains are mostly non-taxable, provokes incentives towards the artificial retention of profits that are not as present in Brazil. This is why, in Chile, the imputation of undistributed profits to the distribution of personal income appears to be necessary when making international comparisons (Atria et al., 2018). This example emphasizes the importance of the DINA project for cross-country comparisons of inequality (Alvaredo et al., 2017).

Figure 1.8: Discrepancy of income across datasets in Chile and Brazil: 2015



Reading: in 2015 the total income declared in tax data in Brazil, which covers 20% of the population represents 49% of national income. The total income in the raw survey represents 58% of national income and 74% in the corrected survey, which are both representative of the entire population. The equivalent income calculated from national accounts represents 85% of national income. Authors' calculations using data from surveys, income tax declarations and national accounts.

Detailed Distribution Table 1.2 depicts a more detailed picture of the impact of our adjustment method on the income distribution of our 5 countries. Again, we take the last available year as an illustration. The first point to note is that the average income of the survey is adjusted upwards every time. The extent of the increase, by definition, depends directly on the shape of the bias that is observed in Figure 1.6. Both the steepness of $\theta(y)$, when it is to the right side of the merging point, and the size of the corrected population (Column 4 in Table 1.1) are decisive factors for the size of such an increase.[17] Figure 1.8 presents the impact of our method on total income. For our two country case studies with the largest corrections to total income, we are able to show that the total income in the corrected surveys is closer to the reference total of "fiscal income" from national accounts. For the cases of Chile and Brazil respectively, our correction bridges about 80% and 60% of the gap between survey income and the reference total from national accounts.

[17] Another way to think about the size of the corrected population is to look at the size of the area between $\theta(y)$ and 1, to the right side of the merging point.

Table 1.2: Income Shares: Raw Survey and Corrected Survey

Raw Survey					
Income groups	**Brazil**	**Chile**	**France**	**Norway**	**UK**
Bottom 50%	16.5%	8.0%	23.4%	25.2%	14.8%
Middle 40%	42.8%	45.2%	47.0%	48.6%	49.6%
Top 10%	40.7%	46.9%	29.6%	26.2%	35.5%
Incl. Top 1%	*11.0%*	*14.3%*	*7.2%*	*5.8%*	*9.4%*
Incl. Top 0.1%	*2.4%*	*3.4%*	*1.5%*	*1.4%*	*2.5%*
Incl. Top 0.01%	*0.6%*	*0.7%*	*0.4%*	*0.3%*	*0.4%*
Incl. Top 0.001%	*0.1%*	*0.2%*	*0.1%*	*0.03%*	*0.04%*
Average income	€8,081	€8,101	€23,367	€37,431	€22,389
Gini	0.53	0.64	0.40	0.37	0.52
Corrected Survey					
Income groups	**Brazil**	**Chile**	**France**	**Norway**	**UK**
Bottom 50%	13.3%	6.6%	23.2%	24.6%	13.9%
Middle 40%	35.2%	39.5%	46.5%	47.7%	46.6%
Top 10%	51.5%	53.9%	30.3%	27.6%	39.6%
Incl. Top 1%	*22.9%*	*16.9%*	*8.2%*	*7.1%*	*13.7%*
Incl. Top 0.1%	*10.5%*	*4.6%*	*2.2%*	*2.2%*	*5.4%*
Incl. Top 0.01%	*5.2%*	*1.3%*	*0.6%*	*0.7%*	*2.1%*
Incl. Top 0.001%	*2.4%*	*0.4%*	*0.2%*	*0.26%*	*0.89%*
Average income	€10,138	€10,949	€23,621	€38,320	€24,081
Gini	0.61	0.69	0.41	0.38	0.55

Notes: The table presents the distribution of pre-tax fiscal income per adult, before the correction and after the correction. Average incomes are expressed in French Euros PPP. Brazil and Chile refer to 2015, while all the European countries refer to 2014.

With respect to income shares across the distribution, the main conclusions that are drawn from the analysis of the Top 1% share in previous paragraphs can be generally extended, with more or less intensity, to other top shares, from the top 10% to the top 0.001% shares. As is to be expected, both the middle 40% and Bottom 50% shares are reduced in all countries. This is consistent with the mechanics of our adjustment, where higher aggregate weight for top fractile incomes must be compensated by a lowering of the amount of middle and lower incomes observed in the population. Again, expectations on the scale of the downward correction of these share can be informed via the size of the bias at the top, as depicted in Figure 1.6. A more general picture of what happens in the whole distribution is presented by the Gini coefficients. In all the latest-year examples, the Gini increases, which reflects a general increase in total estimated inequality using this composite index.[18] The scale of the increase generally reflects the magnitude of the change in income shares.

5 Conclusion

The main objective of this chapter is to provide a rigorous methodological tool that enables researchers to combine income or wealth surveys with administrative data in a simple and consistent manner. We present a new methodology on the combination of such sources, which incorporates a clearer formal understanding of the potential biases at play and a solution to remedy them. The result of our reweighting approach, we argue, should be a more representative dataset that can serve as a basis to study the different dimensions of social inequality. Our algorithm is built in such way that it automatically generates, from raw surveys and tax data, an adjusted micro-dataset including new modified weights and new observations, while preserving the consistency of other pre-existing socio-demographic variables, at both the individual and aggregate level.

This contribution can thus be viewed as an attempt to improve survey representativeness by taking the income distribution into account. While it is common to adjust survey weights in accordance to external information on the distribution of basic socio-demographic variables, our paper motivates the use of auxiliary administrative data sources on the distribution of income, along with other socio-demographic information, to improve the representativeness of the population.

Our procedure has several advantages. First, it is based on an intuitive theoretical framework. Second, our method avoids *a priori* assumptions on the size of the population to be corrected. Instead, it offers a clear procedure to find the merging

[18] Section 4 of Appendix A presents the trends in country Gini coefficients for the full period.

point non-arbitrarily. Third, the algorithm can be applied to a wide variety of countries, both developed and less developed, since it accounts for different levels of data coverage. Fourth, our method respects original individual self-reported profiles and socio-demographic totals for variables other than income. We thus preserve the internal consistency of surveys, while better approximating the external consistency of its income distribution. Although we preserve socio-demographic totals for variables other than income, our method allows for their conditional distribution to vary upon the addition of new income information. However, our method also accommodates the input of distributional information of other variables (age, sex, income type, etc.) if they are available in the tax data. As such, users may also calibrate and correct the survey on covariates of income, in addition to income itself, if reliable statistics exist on their interaction. Ideally, we think that reweighting based on external information on the income distribution could be applied to surveys when employing standard calibration procedures. Finally, it should be clear that this method can serve multiple research objectives – from single-country and cross-country empirical analyses using income statistics as well as their covariates, to research reconciling income and wealth distributions in a national accounting framework, as in the Distributional National Accounts project (Alvaredo et al., 2017).

To the extent of harmonizing our correction procedure among different countries, we stress the importance of analyzing the underlying data in each case. For this, our method provides useful tools to practitioners wishing to assess the population coverage of surveys conditional on income. Figure 1.6 and Table 1.1 are examples of the type of information directly computed by our algorithm. With standard survey and tax data at hand, researchers can perform our correction procedure with relative ease. Given that we make the statistical tools openly available, they could provide the seeds for greater collaboration between national statistics institutes and tax administrations in order to improve nationally representative datasets. The combination of survey and register data is already happening in some countries, with the former gradually becoming anchored to the latter in the most developed cases. National statisticians engaged in the production of surveys could make use of our correction method upon having direct access to data on income and other covariates from government ministries. For many countries in which the majority of the population are not included in income tax statistics or social security contributions, our adjustment could make great gains. For more developed economies, researchers who want to continue to make use of their national household surveys can still do so without concerns over distributive representativeness.

Chapter 2

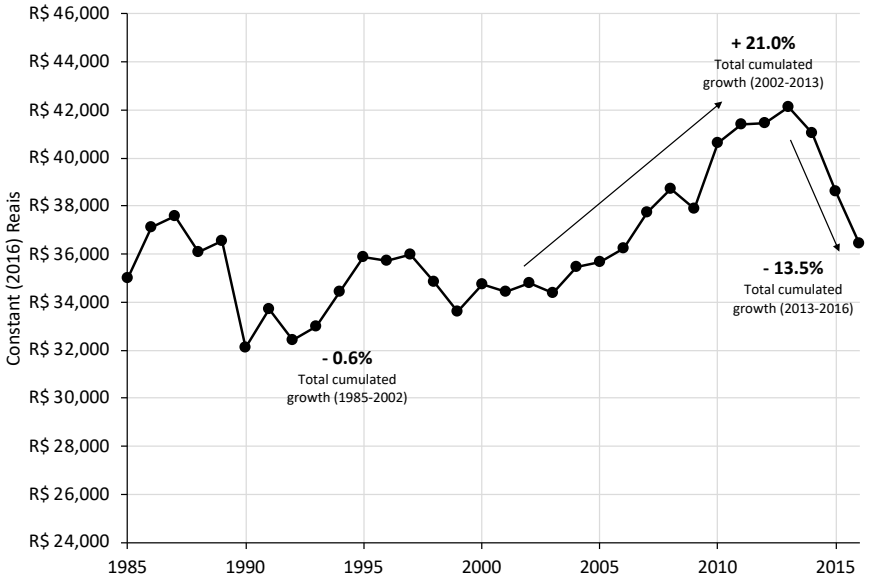
Falling Inequality vs Persistent Concentration: Reconciling Recent Evidence for Brazil from Surveys, Administrative Data and National Accounts

Abstract: This chapter re-examines the evolution of income inequality in Brazil over the 2000s using a novel combination of data sources. I produce a new distributional series of pre-tax national income, reconciling survey data with detailed information from income tax declarations, in a consistent manner with national accounts. I find that inequality within the Bottom 90% of the distribution declined, but concentration at the top persisted at very high levels, which surveys fail to capture. This dichotomy was given by the strong average income growth in both tails of the distribution between between 2002 and 2013. The Bottom 50% share realised higher income growth than the Top 10%, but received an even smaller fraction of growth than the Top 0.1% due to its extremely low command of income. While elites and the poor made gains, the Middle 40% was squeezed, posting less growth than the average for the whole economy. The 2014–2016 recession made everyone's income decline or stagnate, but the relative decline was sharpest for bottom- and middle-income groups. The modest fall in overall inequality came mostly from a compressed labour income distribution, which I document from new estimates combining surveys and fiscal data. However, the policies driving it were insufficient to mitigate the implied inequality of capital income and prevent the growing concentration of national income among elite groups. Partial evidence from tax data suggests that income switching by business owners at the top may have played a role.

1 Introduction

The 2000s are an interesting period in Brazilian history. From an economic perspective, they mark the return of economic growth after at least fifteen years of stagnation since the country's return to democracy (see Figure 2.1). Between 2002 and 2013 average adult incomes expanded by 21% in the world's 8^{th} largest economy and 3^{rd} largest parliamentary democracy. Politically, it is also an interesting period due to the coming to power of the first "labour" government in Brazil since the early 1960s, with the election of Lula da Silva to the presidency and his Workers' Party (PT) compatriots to the Congress in late 2002. New innovations were introduced in the field of social policy to tackle poverty, and a higher share of growing fiscal receipts were dedicated to social spending. The PT government also increased the real value of the minimum wage considerably during their 13-year mandate. This policy shift built the consensus that identified a significant decline in income inequality in Brazil, as elsewhere in Latin America, at least according to official measures relying on household surveys (López-Calva and Lustig, 2010; Lustig et al., 2016).

Figure 2.1: Average National Income Per Adult in Brazil: 1985–2016



Notes: authors' calculations using income and population data from IBGE. Adults are defined as the population aged 20 and above.

From 2014, both the economy and the political system collapsed. The combination of the worst domestic recession since the early 1980s, and a myriad of corruption scandals, created the climate for the impeachment of Dilma Rousseff and her government in mid-2016. It was during this time that a number of studies

began to challenge the accuracy of the survey-based conventional wisdom on income inequality dynamics in Brazil. Medeiros et al. (2015a); Morgan (2015) and Gobetti and Orair (2017b) were among the first to make use of the first public releases of income tax data to show that top income shares and "adjusted" Gini coefficients were in fact relatively stable over the 2000s. These initial findings pointed to the extreme concentration of income in Brazil and the weakness of surveys to reveal its true extent. They also suggested that a harmonised comparison between different data sources was necessary in order to adequately understand the dynamics and discrepancies between Brazil's multiple "income distributions", during a key period in the country's development.

This paper carries out this very exercise. By harmonising and merging various data sources on incomes, including previously unexploited data on labour incomes and formal sector earnings, I offer a more comprehensive evaluation of the evolution of inequality during the 2000s. This contribution is also the first to reconcile distributional data with national accounts in Brazil, following the pioneering research initiated for more developed countries (see Piketty et al. (2018) and Garbinti et al. (2018)).

A number of reasons motivate why we should pursue the combination of survey data, administrative data and national accounts to measure the distribution of income. Most evidently, there has been an over-reliance on household self-reported surveys to assess income inequality. While the "true" distribution is unobserved, household surveys can make an approximation by expanding the frequencies of a representative income-earning sample of the population. The issue with surveys is that they tend not to include complete information on the very rich in the studied country. Despite random and large sampling, income is either not well measured or not observed, due to a mix of sampling error – meaning lack of precision due to limited sample size – and non-sampling error – meaning structurally different response rates across individual types (see chapter 1). The reluctance of rich individuals to disclose the full extent of their income, or to even participate, is an example of the real constraints faced by survey samples. Moreover, statisticians may intentionally remove extreme observations to remove "outliers". Surveys are thus prone to mis-represent the extent of certain income sources in the distribution, compared to what the national accounts reveal. Nevertheless, they been an invaluable source of information for bottom incomes and other socio-demographic characteristics.

Income tax data, on the other hand, better capture the income of affluent individuals, as filing a declaration is obligatory above specified income thresholds, and in many cases (Brazil included), it is accompanied by third-party reporting.

Although not everybody declares income to the tax authorities and some people can be tempted to under-declare their income in order to pay less tax, we can be quite confident in thinking that the people appearing in tax data earn at least what they declare, as they are well identified by fiscal procedures.

The integration of national accounts is important for two main reasons. First, the national accounts are the official source that measures the macroeconomic aggregates in a standardised way across countries. By definition they diffuse information on income growth, but they have been until recently without a framework to distribute the annual proceeds of this growth. The Distributional National Accounts (DINA) framework facilitates the reconciliation of macroeconomic sources of income aggregates with microeconomic distributional sources (Alvaredo et al., 2017). This was not possible in previous studies since the income concept used typically had less of a direct relation with macroeconomic growth, as registered in the national accounts. Second, DINA allow us to see to whom in the hierarchy income originally flows. While it may be argued that "fiscal income" (i.e. income received by households that is subject to assessment by the tax authorities) or "disposable income" (i.e. income individuals actually dispose of after government taxes and transfers) are more relevant to the consumption of individuals, the distribution of national income gives us insights into the distribution of economic resources more generally, among the entire set of income-beneficiaries of the production process. It is thus more intricately connected with the distribution of capital and power. It can also be thought as being more crucial than the distribution of disposable income, as it precedes it in the flow of funds and can thus determine a society's capacity for redistribution.

Similar to the other studies in this literature, I present the main results for the Brazilian case in the form of income shares. These help to stratify the income-generating population into income classes, such that distinct parts of the distribution may be distinguished from each other, as opposed to being confounded in a synthetic indicator, such the *Gini*. Distribution tables depicting income shares are much easier to grasp as their construction is straightforward – the total income of a given fractile in the distribution divided by the total income received by the adult population – and their interpretation is transparent – an income share of 50% for the Top 10% of the distribution gives us a clear sense of how the economic cake is divided.[1] With an index like the Gini we are unable to observe the inequality between the top and the bottom of the hierarchy or between the middle and the bottom or the middle and the top, or within the top itself. More importantly when presenting income levels

[1] A group that receives half of all distributed income, when it only represents one-tenth of the population is a more concrete and visible claim on income concentration than saying that the Gini is 0.60, as the latter is without reference to any particular social group in the hierarchy.

in cash terms (instead of percentages) it makes it possible for people to appreciate their position in the social hierarchy, which can have significant political implications. However, the wide use of the Gini coefficient is at least one reason not to discard it completely. Rather, we should seek to clarify how inequality can be reconciled between different indicators, *as well as* between different data sources. This chapter sheds light on this issue in the Brazilian context, where the fall in the Gini has been a staple input in the inequality debate.

The combination of these different data sources for Brazil confirms the tax-based wisdom that inequality levels are much higher than "officially" estimated by household surveys. Furthermore, there has been less of an overall change in inequality than previously outlined. By contrast, I show that the distribution of labour income and formal salaries did become more equal over the period, which correlates with recent policy initiatives, but does little to compensate the implied inequality of capital income. In any case, we cannot rule out that the decline in top labour income shares is at least partly driven by incentives to shift income from labour to capital sources by affluent individuals over the period.

In the Brazilian case, the distribution of growth can defined by a symmetry between the extreme tails of the distribution, to the detriment of the country's "middle class". In light of recent political events, this finding is not trivial. Overall, total income inequality in Brazil seems to be very resilient to change, at least over the medium run, principally due to inherited economic and fiscal structures that progressive "pink tide" governments have not properly reformed.

The remainder of the chapter is structured as follows. Section 2 presents the data, concepts and the methodology employed to calculate income shares across the entire adult population. Section 3 presents the principal results of the chapter regarding income inequality and growth. Section 4 discusses how we can understand the trends in the context of Brazil's recent history, before concluding. Appendix B provides further methodological information.

2 Concepts, Data and Methods

I exploit three main source of data to estimate inequality across the entire distribution in Brazil: representative household surveys, personal income tax declarations and national accounts. I also make use of administrative data linked to social security contributions to complement the assessment of earnings inequality. To contribute to the debate on the decline of inequality, the quantitative analysis will focus on the years of the 2000s. However, I include estimates from 1995 onwards (after the

stabilisation of inflation) in order to better judge how "different" the early 2000s were.

In the subsequent sections, I present the income concepts utilised in the analysis, the data sources, the adjustments made to them for their harmonisation, and the methodology used to combine them for the estimation of inequality across the different concepts. A more detailed presentation of these sections (for this period) is given in Appendix B.

2.1 Income Concepts

Our aim is to distribute net national income, as it appears in the national accounts. This is the total income that flows annually to national residents of Brazil (comprised of various institutional sectors – households, non-profit institutions, corporations and the government), after accounting for depreciation of fixed capital. The DINA procedure seeks to distribute this total among adult individuals aged 20 years and over residing in domestic households. Thus, the inequality estimates we report are of the distribution of per adult income. We take the income of married couples as being equally divisible between the comprising adults. The equal-splitting of couple income has the advantage of not "overestimating" inequality by not underestimating the resources available to non-working spouses, especially in societies with relatively low female participation in the labour market.[2]

Using survey data, we can first estimate a distributional series of "survey income". This income concept is constructed to be as equivalent as possible to the one that is assessed by the tax authorities from the personal income tax declarations, which are taken to be the benchmark for incomes at the top end of the distribution. In sum, this includes wages and salaries, social insurance benefits (gross of contributions), self-employment income, interests, rents, distributed business profits and dividends, and capital gains. Using tax data to correct for top incomes in the surveys, we arrive at a series of "fiscal income". This can be interpreted as the total income that would be reported by individuals for fiscal purposes, as given by values declared on the survey questionnaires and tax returns. It thus corresponds to *pre-tax post-replacement fiscal income*, i.e. income received by individuals before personal income taxes, employee and self-employed social contributions (but after employer contributions), and legal

[2]This perspective assumes that couples redistribute income between their members, as if all couples have joint accounts with equal access to the resources. This assumes away any unequal bargaining power among couples with unequal income flows, which may be an overly optimistic treatment of intra-household allocation of income. But the assumption of no sharing of resources is unrealistic too. I judge it is preferable, where data does not allow for more refined calculations, to be on the lower bound of the inequality estimate (assuming equal splitting) rather than on an upper bound (assuming zero sharing of income). See Alvaredo et al. (2017).

deductions, but after accounting for social security benefits in cash (unemployment insurance and social security pensions).

"Fiscal income" is distinguishable from "national income" insofar as it only concerns distributed income received by physical persons that is assessed by the tax office for fiscal purposes. It should also be distinguished from "taxable income", which is the income that is ultimately taxed after the application of legal deductions. Some components of income may have to be declared on the tax returns but are not ultimately "taxable".[3] Pre-tax fiscal income constructed in this way can also be related to an equivalent total from national accounts.

The next stage involves moving from fiscal income to "personal income". This implies that we factor in flows of income appearing in national accounts that (1) get attributed to households but are not included in fiscal income, such as imputed rents, investment income attributable to insurance and pension funds and social contributions made by employees and self-employed workers; and (2) are not attributed to households, but rather to corporations, such as undistributed corporate profits (i.e. the net primary income of corporations). Finally, moving from personal income to "national income" means that we must account for incomes that flow to the government rather than to households, namely government net capital income (government share of undistributed corporate profits, net interests and other incomes) and net production taxes. More precisely, total pre-tax national income can be computed as follows:

Total pre-tax fiscal income[4]
− Social contributions (D61, S14)
+ Imputed rent for owner-occupiers
+ Investment income attributable to insurance policyholders (D441, S14)
+ Investment income payable to pension entitlements (D442, S14)
+ Household/NPISH component of pre-tax undistributed corporate profits (B5n, S11+S12)
= Total pre-tax personal income
+ Government factor income[5]
+ Pension and other social insurance surplus[6]

[3]This may vary with countries. In the case of Brazil, as we have presented, it is explicit, as the tax declarations include a section for declaring non-taxable incomes.

[4]This can vary by country, but usually involves the sum of salaries + gross operating surplus + gross mixed income + gross interests + dividends + social security cash benefits + gross current transfers - imputed rent - capital depreciation of households.

[5]This equals net property income received by the government (D4n, S13).

[6]This equals pension and other social contributions (D61, S14) − pension and other social

+ Net production taxes received by the government (D2-D3, S13)
= **Total pre-tax national income (DINA)**

It may seem questionable to impute income flows to individuals, which are not directly received during the year. A case in point is the undistributed profits of domestically-owned corporations. Under conventional practice, such retained earnings in "legal persons" (i.e. corporations) would rarely be considered individual or family income. Yet, these income flows must ultimately have beneficial owners, which either exercise direct control over the flows or have future entitlement to them. But more importantly, excluding these "missing" flows introduces arbitrary biases in distributional series between countries, which are the result of between-country differences in legislation. In addition these biases can be time-varying. In the example of corporate profits, the extent to which they are distributed or not may well depend on incentives derived from country-specific tax laws or corporate governance legislation, which can change over time. And these decisions have important effects for the inequality we wish to measure.

Distributing national income gets around these biases. However, an important part of my analysis will be the comparison of the national income distribution with the distribution of other income concepts, namely survey income (as reported in the "uncorrected" surveys) and fiscal income (after correcting surveys with tax data). The discrepancies between the three series can have different implications for economic analysis and public policy.[7]

2.2 Survey Data

The survey data I use is from the *Pesquisa Nacional por Amostra de Domicílios* (PNAD), the large, nationally representative household survey organised by the IBGE (Brazil's National Statistics Bureau). The survey runs annually except in the years coinciding with the National Census (approximately once per decade). It consists of a household wave and an individual wave, the latter's sample being approximately 350,000 people per year. Moreover, the survey is weighted by the population census.

insurance benefits (D621+D622, S14).

[7] While much economic analysis is interested in individual or household monetary "welfare" (either from gross or disposable income), questions of economic *security*, *capability* or even *efficiency*, more generally, should not be overlooked. What may apply to one individual or a group of individuals may not apply in the aggregate. Ultimately, the pass-through of incomes from the pre-tax national distribution to the pockets of individuals is mediated institutionally, by corporate legislation, fiscal policy and social conventions, etc. To better understand the entire process it is necessary to begin with the pre-tax national distribution. This conceptual analysis can be interpreted as bridging the gap between the micro focus on individual welfare and the macro focus on economic aggregates.

I use the individual-level micro-files for the PNAD between 1995 and 2015, which are publicly available on the IBGE's website.[8]

The data are nationally representative with the exception of the waves before 2004, which exclude the rural areas of six northern states (Rondônia, Acre, Amazonas, Roraima, Pará and Amapá). According to the Censuses and recent PNADs the share of the excluded population corresponds to between 1.5% and 2% of the total. I do not make any adjustments to the data to compensate for this, given that the impact of the excluded population on income shares is largely imperceptible in the 2004 PNAD. The survey reports individuals' pre-tax monthly incomes in the reference month (usually September) by source, a long with a host of other socio-demographic characteristics. It has been the IBGE's flagship household survey for decades and the principal source of income data for research on poverty and inequality in Brazil, despite its common limitations.[9]

2.3 Administrative Data

The second broad source of incomes comes from administrative data. The main source I exploit here are tabulations of personal income tax declarations (DIRPF). These are published directly by the Brazilian Federal Tax Agency (for the years 1996-1998, 2002, 2007-2016), or reproduced from books/research papers (for 2000 and 2006), when years from official sources were unavailable. Unfortunately, the Tax Agency has never granted external researchers access to the personal micro-files. Nonetheless, the published tabulations are detailed enough to be able to produce accurate estimates of the top end of the distribution, under the assumption that tax data contains the most credible information incomes above a given threshold. The tabulations include total number of declarations by intervals of total assessed income, deductions and the net tax base.

A useful feature of these tabulations is that they decompose total assessed income into three "fiscal" categories per bracket: "gross taxable income", "exclusively taxed income" and "non-taxable income", such that we observe the total personal income of declarers and not just that which is strictly taxed. Taxable income is the portion of declared income subject to the progressive income tax schedule after the application of deductions. It comprises mainly of labour income (wages of salaried and self-employed workers, the taxable portion of pensions), as well as property rent and royalties. Exclusively taxed income includes categories of income withheld at source and taxed

[8] https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html?=&t=downloads.

[9] See section 2 of Appendix B for further details.

according to a separate schedule. These mainly concern capital income (other than rents and royalties), such as capital gains and interests from financial investments, but also labour incomes such as the 13^{th} salary (i.e. the annual obligatory bonus). Over the 2007-2016 period, these incomes have accounted for about 10% of total assessed income. Lastly, non-taxable income refers to income exempt from the personal income tax. These include a host of labour income and social benefits, such as compensation for laid-off workers, the exempt portion of pension income for over 65s, the exempt portion of agricultural income and scholarships, among other items, and capital incomes such as distributed profits and dividends of all incorporated businesses and small unincorporated businesses, interests from savings accounts/mortgage notes, etc. Profits and dividends are by far the largest component of this category, covering over one-third of the total. In the aggregate, these exempt incomes represent almost 30% of total assessed income.

All in all, we avail of between 8.7 million tax returns in 1996 (about 8% of the adult population) and 28 million returns in 2016 (about 20% of the adult population). This significant increase is both due to economic expansion and the continuous modernisation of the tax agency.[10] A complex feature of the tax statistics is, that for numerous years, they present tabulations for distinct income concepts to the one we are interested in exploiting from them, namely total pre-tax fiscal income (as highlighted in Table B.1 of Appendix B). Not only does the concept of reported income change across years, but so does the concept of the ranking income across brackets. While for the 2006-2016 period, tabulations present total income ranked by total income, for prior years (1996-1998, 2000, 2002) we only have tabulations ranking gross taxable income by gross taxable income. I explain below how total incomes are estimated from these earlier tabulations.

When using tax data, a valid concern is the presence of evasion. In the Brazilian case this is no different. However, the design of the country's personal income tax system merits some consideration in this context. Firstly, since some important components of capital income are exempt from the personal income tax, such as dividends, this reduces the incentives to under-declare these types of income. For instance, when comparing the dividends declared in the tax statistics with those in national accounts one finds that the difference is around 3% on average. Moreover, capital income in the form of capital gains and interests from financial investments are withheld at source and taxed exclusively either at flat rates or at rates depending on the nature and maturity of the investments. This is facilitated by specific monitoring

[10] See Figure B.1 of Appendix B for a historical perspective on the number of returns and taxpayers.

programs used by the federal tax office, which match declared personal incomes from tax records with financial information provided by banks, through the *Declaração de Informações sobre Movimentação Financeira* (DIMOF).[11] Nevertheless, a certain amount of measurement error in the declaration of income should be expected, as well as the possibility of other income sources (typically property rent or self-employment income) to be under-declared.[12]

In addition to income tax declarations, I also make use of publicly available administrative data on employee earnings provided by the National Institute for Social Security and the Agency for Social Insurance within the Federal Ministry of Finance.[13] The data are tabulations of earnings (wages and salaries) of all employees in the formal sector who contribute to the general regime of social security (RGPS). Specifically this category of workers includes contributors on a full-time or temporary formal contract, employed in the private or public sector. It excludes public servants, who contribute to their own regime of social security (RPPS). It also excludes domestic employees ("empregado doméstico"), self-employed workers ("contribuinte individual"), contractual workers ("trabalhador avulso"), rural producers and artisans ("segurado especial") and any voluntary contributors ("segurados facultativos").

These tabulations contain intervals of earning thresholds (defined as multiples of the minimum wage), alongside the number of contributors and the total value of earnings per interval for for the period 1996-2016.[14] On average, these contributors represent about 32% of the adult population, and about 56% of the employed population. Reflecting the expansion of the economy and formal employment over the period, the number of employee contributors increased considerably, from 21.6

[11] The DIMOF is an obligatory declaration by banks (including credit cooperatives and savings and loan associations), through which information is passed on to the government about all financial operations undertaken by the banks' clients. It was initiated in 2008. Prior to 2008 the government could avail of the financial transactions tax (the *Contribuição Provisória sobre Movimentação Financeira*, in place between 1997 and 2007) to crosscheck the information about financial investments provided by contributors. All individuals are required to provide their bank account details on their declarations.

[12] The under-declaration of self-employment income may not be as large as expected for two reasons. First the DIMOF program applies to all workers, independently of the nature of their occupation. Independent workers would have to carry out all of their transactions in cash for them to avoid a bank trace. Second most own-account workers, on the basis of anecdotal evidence, create a legal business under their name and register their income as profit withdrawals or dividends so that they appear on the declarations but avoid paying the income tax.

[13] http://www3.dataprev.gov.br/infologo/inicio.htm.

[14] Although the contributing salary of earners in Brazil is capped at a given, annually adjusted, threshold, the data presents the total remuneration of employees per bracket, whose highest values are well above the contributing threshold. For instance, in 2016 the contributing ceiling was set at a value of BRL 5,189.82, which corresponds to just under 6 monthly minimum wages. By contrast the top bracket in the 2016 tabulation is defined by values exceeding 40 monthly minimum wages. Thus we can be sure that top-coding is not an issue.

million individuals in 1995 to 51.3 million in 2016.

2.4 National Accounts

The last source I exploit are the national accounts of Brazil.[15] The integrated national accounts (*Contas Econômicas Integradas* (CEI)) are available from the IBGE for the years 1995-2015 (IBGE (2000; 2017). For 2016 we only possess totals of the main aggregates, so we extrapolate totals for all sub-components maintaining the share of each component in national income for this latter year. The CEI follow the United Nations classification of institutional sectors and variables (SNA-93 for 1995-1999 and SNA-08 from 2000 onwards). All variables we use are sourced from the CEI, except for values of imputed rents, which we take from the IBGE's *Tabelas de Recursos e Usos* (TRU), and estimates of fixed capital consumption, which I source from calculations made by Morandi and Reis (2004) and extended by IPEA over the 1950-2008 period (about 11% of gross national income on average).[16] In order to obtain fixed capital consumption for the household sector, I assume that it represents 25% of total CFC for the national economy, which is in-line with the ratios of Chile and Mexico, the only two Latin American countries with this data available on `wid.world`.[17]

2.5 Data Adjustments

In order to adequately merge our distributional data, we need to account for the different degrees of coverage in each source. Thus, we have to harmonise tax and survey data, so that they adhere to the same income concept before combining them with additional information from national accounts.

2.5.1 Tax Data

I make three adjustments to the tax data. The first concerns the subtraction of fixed capital consumption of unincorporated businesses. I use published data on an allowance for the business expenses of independent workers required to keep accountancy books (e.g. doctors, dentists, psychologists, lawyers, independent commercial agents, etc.). These expenses are can be identified in the deduction

[15] Section 3 of Appendix B explains the historical evolution of national accounting in Brazil.

[16] See `http://www.ipeadata.gov.br/Default.aspx`. For 2009-2016, I assume the same proportion of CFC in GNI as 2008 (around 13%).

[17] Government CFC and NPISH CFC are both set equal to the gross operating surplus of each sector in the national accounts, which amounts to about 11% of total CFC for the former and about 1% for the latter. Thus, we estimate corporate CFC as a residual, which comes to a share of about 63% of total CFC.

"livro caixa" in the tabulations, which I use to subtract from total assessed income, whenever this information is available. For years without such information, I impute this deduction by applying bracket-specific percentages, based on the closest years with complete information.[18] Such expenses are not identifiable in the household survey, but we know from the tax statistics that these are generally concentrated among higher incomes, for which the tax data are more reliable.

Second, tabulations for the years 1996-1998, 2000 and 2002 present information only on "gross taxable incomes", from which total incomes need to be estimated. Again, we rely on the closest neighbouring years with complete information on gross taxable incomes and total incomes (i.e. years 2007 to 2016) to impute the latter to these early years. Section 5.4 of Appendix B describes in detail the imputation method employed.

Third, given the random nature of the tax unit (for married couples), and our interest in distributing income individually, we need to distinguish declarations made by single individuals from declarations made by couples. This is not all that straightforward from the information in the tabulations. Since joint filing by couples is voluntary, this brings forth the complication that not all single declarations are made by persons whose civil status is actually single. The breakdown of declarations by civil status is only available for historical years (the 1970s and 1980s). Thus, for the more recent period, I have to rely on partial evidence. Specifically, I make use information from the deduction for dependants available in the 2007–2016 tabulations, to calculate the share of joint declarations per bracket. For earlier years, I estimate an average per bracket share of single declarations of the closest neighbouring years. Section 4.2 in Appendix B describes the estimation process in detail.[19]

2.5.2 Survey Data

In order to reconcile incomes from the surveys with those from tax data, we need to make four types of adjustments.

First, we need to impute certain types of income that have to be reported on tax returns, but are not reported in the surveys due to their infrequent nature. These incomes correspond mostly to benefits of formal employees defined by labour laws, such as unemployment insurance, FGTS, *abono salarial*, the holiday bonus and the

[18] See section 4.1 of Appendix B for further details.

[19] This estimation method assumes that married individuals filing separately are either single or married to other individuals whose income falls in the same bracket, which may not be true. This means that we may over-state inequality compared to the perfect equal-split case (where the total income of actually existing couples is divided by 2), and may under-state inequality as compared to the pure individualistic case (where each spouse is assigned his or her own income). This limitation can only be remedied by having access to the Brazilian income tax declaration micro-files.

13^{th} salary. Fortunately, these benefits can be easily calculated by applying labour laws to standard PNAD variables such as employment status (see section 4.3 in Appendix B for details).

Second, we have to reclassify incomes to three main "fiscal income" categories: labour, capital and mixed income. Labour incomes comprise the labour market earnings of employees, pensions and survivors' benefits, a fraction of employers' earnings, and all imputed and/or observed labour-related benefits. In my benchmark scenario, the labour fraction of employers' earnings is equal to income up to the exempted threshold defined by the personal income tax. Excess incomes above that threshold are classified as "capital", alongside financial (that is, "other incomes" greater than the current minimum wage) and rental incomes. In turn, mixed incomes are the earnings from the main occupation for self-employed individuals with no employees.

Third, I estimate "annualisation factors" to take into account both monthly inflation rates and the reference months of the survey (typically, September). As expected, all multipliers after the macroeconomic stabilisation in 1995 are very close to 12.

Finally, I keep only adults aged 20 years and above and discard individuals with missing total income.

2.6 Comparison of Raw Totals Across Sources

Table 2.2 shows the best approximation that can be made of the equivalent income totals from our three main data sources for the 2000-2015 – the years for which the data is most detailed. The comparisons confirm that the surveys severely underestimate capital incomes, while they do a much better job at capturing labour incomes. Despite its restricted population, the tax data is better equipped to capture a higher fraction of capital incomes, but by only covering 20% of adults, it does less well in capturing labour incomes as compared to the surveys. This reflects the concentration of capital income with respect to labour income – almost half of all labour incomes registered in national accounts flows to non-filers i.e. the bottom 80% approximately. It must be stated that some measurement error is expected when computing the income totals across the three sources, especially from tax data, such that certain components may be over/under-estimated. Only greater transparency from the tax office will improve the accuracy of these estimates.

Figure 2.2: Comparison of Incomes Across Data Sources: 2000–2015

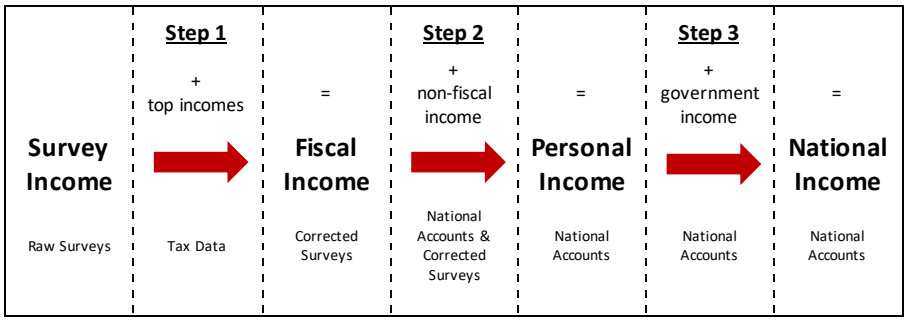
Year	Total income (% national income)				Incl. labour income (% national income)				Incl. capital income (% national income)			
	SNA National Income	SNA Fiscal Income	PNAD	DIRPF	**SNA National Labour Income**	SNA Fiscal Labour Income	PNAD Labour Income	DIRPF Labour Income	**SNA National Capital Income**	SNA Fiscal Capital Income	PNAD Capital Income	DIRPF Capital Income
2000	**100%**	80.8%			**64.5%**	61.9%			**35.5%**	18.9%		
2001	**100%**	81.3%	57.6%		**64.8%**	62.3%	50.1%		**35.2%**	19.0%	7.5%	
2002	**100%**	79.5%	57.4%		**63.6%**	62.6%	50.1%		**36.4%**	16.9%	7.4%	
2003	**100%**	77.4%	55.5%		**63.3%**	62.0%	48.5%		**36.7%**	15.4%	6.9%	
2004	**100%**	75.1%	54.6%		**62.9%**	60.3%	47.9%		**37.1%**	14.9%	6.7%	
2005	**100%**	76.8%	56.9%		**63.7%**	61.4%	49.9%		**36.3%**	15.4%	7.0%	
2006	**100%**	77.0%	58.2%		**63.7%**	61.8%	51.0%		**36.3%**	15.2%	7.2%	
2007	**100%**	76.0%	55.3%	41.7%	**63.4%**	60.8%	49.1%	33.1%	**36.6%**	15.2%	6.1%	8.6%
2008	**100%**	76.5%	54.8%	46.1%	**64.0%**	60.1%	48.6%	35.0%	**36.0%**	16.5%	6.2%	11.1%
2009	**100%**	79.8%	55.9%	45.6%	**65.3%**	61.7%	50.1%	35.3%	**34.7%**	18.0%	5.9%	10.3%
2010	**100%**	78.7%		44.6%	**64.8%**	60.3%		33.8%	**35.2%**	18.3%		10.9%
2011	**100%**	80.1%	51.4%	46.2%	**64.8%**	60.5%	46.2%	34.0%	**35.2%**	19.6%	5.2%	12.2%
2012	**100%**	80.1%	53.9%	45.9%	**65.4%**	61.2%	48.0%	34.4%	**34.6%**	18.9%	5.9%	11.5%
2013	**100%**	79.5%	53.3%	45.1%	**65.2%**	61.3%	47.6%	33.6%	**34.8%**	18.1%	5.7%	11.5%
2014	**100%**	81.7%	54.9%	46.6%	**65.6%**	62.9%	49.0%	34.9%	**34.4%**	18.9%	5.9%	11.8%
2015	**100%**	85.2%	55.4%	48.5%	**66.7%**	64.9%	50.0%	36.1%	**33.3%**	20.3%	5.5%	12.4%

Notes: "SNA" refers to the "System of National Accounts"; "PNAD" is the household survey; and "DIRPF" are the personal income tax declarations. The table shows the ratio of the income of each dataset to the total net national income of the economy. The difference between "SNA National Labour Income" and "SNA Fiscal Labour Income" is that the former includes pension benefits and deducts corresponding contributions, while the latter only includes pension benefits to be consistent with the definition of income in tax data. Interpretation: in 2015, the total income we measure in the tax data accounts for 49% of national income, while the equivalent income concept from the survey is 55%. The equivalent income concept estimated from national accounts is 85% of national income. While 66% of total national income is labour income, 65% is fiscal labour income (received by households), 50% is labour income in the survey and 36% is labour income in the tax data. The PNAD incomes are from the microfiles provided by the IBGE, while incomes from the DIRPF are from detailed tabulations provided by the Secretaria da Receita Federal do Brasil. SNA data is from IBGE. Percentages may not add up due to rounding. Mixed income is divided up between labour (70%) and capital income (30%).

2.7 Methods to Estimate DINA: 1995-2016

Figure 2.3 summarises our estimation procedure for DINA into three broad steps. Each step defines the income concept whose distribution we are interested in.[20]

Figure 2.3: Framework for Building Pre-Tax DINA for Brazil



Notes: authors' elaboration. Each income concept is associated to an underlying data source for its construction. See text for definitions.

2.7.1 From Survey Income to Fiscal Income

Step 1. In the first step, the transition from "survey income" to "fiscal income" is made by correcting the former for income under-coverage at the top using tax data. To do so we estimate an income distribution from the surveys and the tax data. The former is straightforward, given that we avail of survey microdata. We distribute all income that is defined as assessable by the Tax Agency among "equal-split adults". This concept equally splits the income of married couples, taking the income of single adults as given. Concerning the tax data, we use the generalized Pareto interpolation (gpinter) method of Blanchet et al. (2017) to turn the tabulated distribution into a continuous distribution for the portion of the population included in the tabulation. Furthermore, we individualise the distribution of tax income, using the share of single declarations estimated from the tabulations (see section 2.5.1). The assumption involves equally splitting the income of married couples filing jointly.[21] We then compare income levels in the interpolated tax distribution to those in the survey distribution to verify the extent of income under-coverage in the latter and incorporate the tax incomes into the survey data using a novel combination method developed in chapter 1. Using the distribution of incomes from the tax data, this method reweights survey observations within the original support (to address non-sampling error), and then replaces observations at the top by a distribution drawn from the tax data (to

[20] For greater details of these steps see Appendix B.

[21] These interpolations are performed using the program available at `wid.world/gpinter/`.

address sampling error), matching the survey covariates to it by preserving the rank of each observation. This method confirms that substantial income under-coverage occurs at the top of the distribution in Brazil.[22] At the end of the procedure we have a distributional series for total fiscal income for the overlapping years of both datasets. For years where only one of the two datasets is available we proceed to extrapolate/interpolate according to information from the closest relevant years for which both sources exist.[23]

2.7.2 From Fiscal Income to Personal Income

Step 2. In the second step, we incorporate into the fiscal income distribution all private non-fiscal incomes that can be attributed to the personal (household) sector, in order to arrive at a distributional series for pre-tax post-replacement personal income.[24] This requires the identification and imputation of missing personal capital income as well as to the imputation of social contributions, which are not included in the gross income assessed for fiscal purposes. The missing personal capital income is income attributed to households but not declared to the tax authorities, and also income that does not get attributed to individuals, but rather to corporations. The first part we can identify as investment income attributable to pension and insurance funds held by individuals and imputed rents, while the latter are the undistributed profits of privately owned corporations associate to Brazilian residents. Thus, in this step we impute the following items, whose totals are obtained directly from national accounts (IBGE 2000; 2017)[25]:

Social insurance contributions (D61, S14)
Imputed rent for owner-occupiers
Investment income attributable to insurance policyholders (D441, S14)
Investment income payable to pension entitlements (D442, S14)
Household/NPISH component of pre-tax undistributed corporate profits (B5n, S11+S12)

Overall, social contributions make up about 14% of national income over 1995-2015; imputed rent makes up about 9%; investment income on insurance/pension funds account for 1%; and the household share of undistributed profits represent about 6% of national income.[26]

[22] See Figure B.10 in Appendix B for evidence on selected years.

[23] See Table B.2 in Appendix B for an overview of the strategy we employ for each year where we make these imputations for the estimation procedure of the 1976-2016 period.

[24] We include the non-profit sector (NPISH) inside the personal sector.

[25] Full details of the sources and estimation procedure of the totals and their distribution are presented in Section 6.2 of Appendix B.

[26] The household share of undistributed profits corresponds to the net primary income of the

At this stage, we have the aggregates for each of the non-fiscal income categories we require in the estimation procedure. Thus, the second ingredient we need is to impute their distribution among Brazilian individuals. We perform this imputation for all four categories using the corrected PNAD survey. Social contributions are imputed analogously to the imputation of labour benefits per adult (see section 2.5.2), by applying prevailing rates to survey variables, such as earnings and employment status, with the aid of a few simplifying assumptions. For imputed rents, we estimate a per adult value for each household using information on rental values and dwelling characteristics in the household wave of the survey, which we then merge to the individual wave of the survey.[27]

For fund income, we distribute the aggregated total (D441 + D442) by assuming that it follows the distribution of wages for earners who contribute to a social insurance fund (i.e. who have positive social contributions). For undistributed corporate profits we impute their distribution in line with the joint distribution of financial income (interests and dividends) and employer profit withdrawals.[28] We test seven other imputation "scenarios" for undistributed profits based on the type of income used for the imputation and the estimation of employer capital withdrawals (see sections 6.2 and 6.4 of Appendix B for further details).

The different ways of allocating undistributed corporate profits among the population each reflect expectations we make on the 'beneficial owners' of them. Our preference is to base our imputation on the remuneration of employers (business owners) that can be attributed to profit withdrawals and the distribution of financial income, which is likely to be the best correlate of owning enterprises. Hence, our benchmark scenario is justified by what we deem to be the most reasonable allocation of undistributed profits, given the variables at our disposal and our knowledge of the tax and social contribution systems.[29]

corporate sector (B5n) attributable to the domestic household sector (as opposed to the government or foreign components), through their ownership of equity shares. We estimate these using the household sector share of total corporate equity liabilities (AF5) in the financial accounts for 2010-2015 (IBGE, 2017). For 1995-1999 estimates, we linearly interpolate between the 2010 estimate and an estimate for 1985 drawn from a business survey conducted by *Visão* magazine on the ownership structure (between private domestic, public and foreign sectors) of the 8,094 largest incorporated firms in the country. See Section 6.2 of Appendix B for further details.

[27] See Section 4.3 of Appendix B for more details

[28] The latter are estimated as the portion of employer income that exceeds the exemption threshold for the personal income tax. The survey only reports the total income of employers, which conflates their wage and their profit withdrawals.

[29] This choice is further validated by the sensitivity analysis we depict in section 6.4 of Appendix B. For the other non-fiscal income imputations, we do not test various scenarios for two main reasons. First, multiple scenarios can be less easily constructed given the auxiliary variables we use for them. And second, the assumptions we make on their distribution seem to be a reasonable first approximation, given the incomes we are dealing with. The imputation of undistributed profits allocates the quasi-totality of them to the Top 10% in the distribution (about 95% of the total generally), and the Top 1% (about 70%). By contrast, our imputations for insurance/pension fund investment income allocates around 50% of the total to the Top 10%; social contributions are almost split about 50-50 between the Top 10% and the Middle 40%; while generally about one third of imputed rent is allocated to the Top 10%, Middle 40% and Bottom 50%.

2.7.3 From Personal Income to National Income

Step 3. Up until this step, we have imputed all the private non-fiscal incomes to the personal sector. The final step involves imputing the remaining categories of income to arrive at a national income distributional series. Here we must account for:

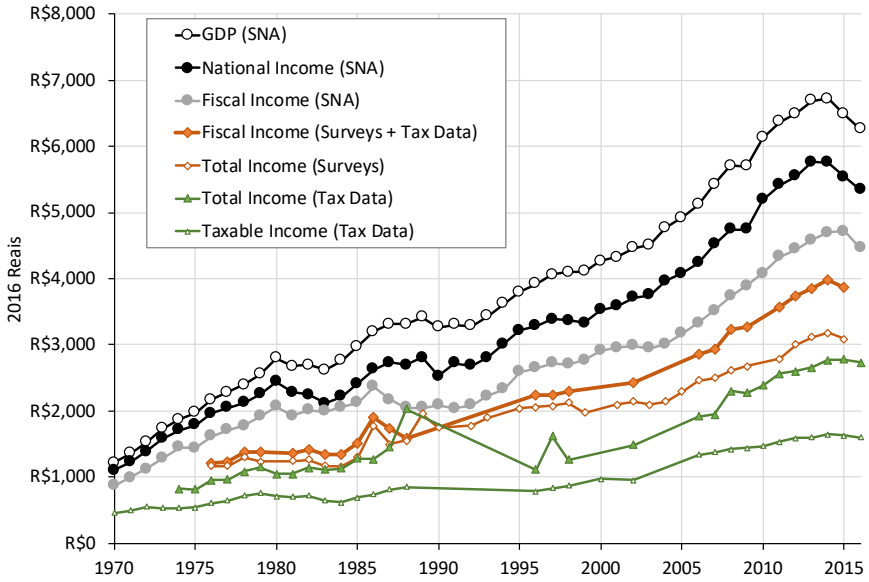
Government factor (capital) income (D4n, S13)
Net production taxes received by the government (D2-D3, S13)
Pension and other social insurance surplus (D61 – (D621+D622), S14)

I assume the distribution of these three categories to be proportional to the distributional of personal income, given the lack of a more precise method to impute each item. Concerning the third category, which has registered a slight deficit in Brazil over the last 20 years, one could argue that the deficit is covered by taxes/levies applied across the entire population on different sources of income/consumption. Likewise for a surplus, which should be attributed proportional to contributing income.

Net production taxes received by the government include value-added taxes and other product taxes, net of product subsidies, (D21-D31) and other taxes on production, net of subsidies (D29-D39). Along with government net property incomes, they form part of government's net primary income. In theory, these taxes should be imputed in relation to their incidence, which can be approximated by consumption, except for property taxes in D29, which should be imputed in proportion to housing wealth or imputed rent (Alvaredo et al., 2017). However, in practice imputations based on tax incidence are difficult to make without access to good data. Moreover, imputing these taxes in proportion to total personal income seams a reasonable first approximation, given that they are based on consumption-type taxes and to production/property-type taxes, which practically cover the entire population. Thus, in proceeding this way for these remaining items, we only seek to normalise the distribution to national income. This step, by definition has no impact on inequality, such that the distribution of personal income is equal to the distribution of national income.

Figure 2.4 depicts the evolution of the main aggregate income concepts that we mobilisee in the estimation process depicted in Figure 2.3 from the survey data, the tax data and the system of national accounts (SNA), as compared to GDP. Through the estimation process just described, we arrive at the distribution of national income, which grew at 1.5% per year on average. It can be seen that other income aggregates grew at broadly comparable rates. Total (pre-tax) income in the surveys represent about 55% of national income, and approximately 70% of an equivalent income total from SNA. Total pre-tax income from tax data, while only covering 20% of the population in surveys, represents close to 50% of national income, and close to 60% of total fiscal income from SNA. The merged tax+survey fiscal income series covers over 80% of an equivalent total from SNA

Figure 2.4: Evolution of Income Aggregates by Income Source in Brazil: 1995–2016



Notes: authors' calculations based on data from surveys, tax tabulations and national accounts (SNA). Nominal values are deflated using the GDP deflator. National income covers about 85% of GDP. The remaining 15% is due to capital depreciation and net foreign income. Fiscal income (SNA), covering an average of 80% of national income, represents an equivalent income total to the income declared on tax returns, which itself accounts for close to 50% of national income. The equivalent income total in the raw surveys covers 55% of national income, while our corrected survey series covers close to 70%. This series merges survey income with assessed incomes from tax data, of which about 60% comprises of taxable income and 40% of non-taxable income.

and equals an average of 70% of total national income over the period. Thus, using this aggregate as a benchmark, approximately 30% of national income is accounted for by "non-fiscal income", which is imputed as previously outlined.

2.8 Estimating the Distribution of Labour Income

To estimate the distribution of labour income I combine surveys and tax data, following the procedure described in section 2.7.1, except that I restrict the income concept to the measurement of pre-tax labour income rather than total income. This concept takes into account all labour income from salaried and independent work, including labour benefits (pensions, unemployment insurance, annual bonuses and other top-ups), net of social security contributions. This is a straightforward task with the survey microdata (see section 2.5.2). However,the tax tabulations present more of a challenge, as there is no clean decomposition of total income between labour income, mixed income and capital income. I use the tabulations of "taxable income" (salaries, pensions, self-employed labour income, and property rent) and the generalized Pareto interpolation (gpinter) method in order to

interpolate the continuous distribution over 127 percentiles. I then merge this distribution into the survey, using the method of chapter 1 and the same income concept of "taxable income" in both sources. Using the adjusted survey microdata, I exclude property rent and social contributions from taxable income and add missing labour income components (annual bonuses, low-income top ups – *abono salarial* – and unemployment insurance) to arrive at a corrected series for pre-tax labour income.[30]

To complement this analysis, I also make novel estimations of the distribution of formal employee earnings from the private sector using social security data (see section 2.3). To do so I apply the gpinter to the tabulation of formal earnings. I do not make a combination here with data on earnings for similar workers observed in the survey, given that the coverage of formal salaries in the fiscal data is superior to that of the surveys – formal employees in the private sector contributing to social security in the survey represent between 65% and 75% of the number appearing in the fiscal tabulation (with the share declining over time). This tendency further reveals the extent to which surveys underrepresent income earners for the income concept they are deemed to be most reliable at reporting.

The pre-tax labour income that we distribute using surveys and tax data corresponds to a total of close to 70% of total national labour income as measured from national accounts. The remaining 30% comprises of missing employer fringe benefits, payroll taxes, employer social contributions, as well as any tax evasion. In contrast, the formal earnings of employees that we distribute account for about 35% of national labour income, and about 45% of wages and salaries registered in national accounts. 'while our series on formal earnings can be seen a subset of the more complete labour income series, it is built on pure administrative data.

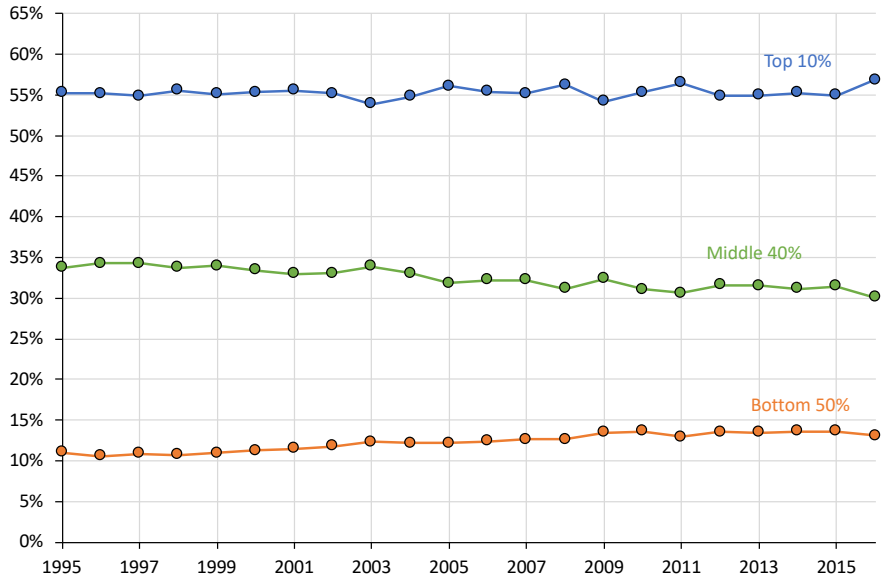
3 Inequality and Growth in Brazil: 1995–2016

3.1 Total Income Inequality

Figure 2.5 presents our corrected estimates for the full distribution of national income in Brazil, separating the adult population into the Top 10%, Middle 40% and Bottom 50%. The first finding to highlight is the extent of income concentration in Brazil in recent years. The richest 10% in the population receive at lease 55% of total income, while the bottom half in the population, a group five times larger, receives less than 15%. The Middle 40% in the distribution receives less than one third of total income. This reveals that inequality in Brazil is about the large division between the top and the rest of the income hierarchy. Second, the trends over the twenty-year period show a resilient concentration at the top and a compression of the distribution within the bottom 90%. Despite the gains made by the Bottom 50%, which increased its share of national income from 10.9% to

[30]The missing labour income components are imputed into the raw survey prior to its fusion with tax data incomes. See section 2 in Appendix B.

Figure 2.5: Shares of National Income in Brazil



Notes: Distribution of pretax national income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining survey, fiscal and national accounts data). Equal-split-adults series (income of married couples divided by two).

13.1%, the Top 10% income share also evolved positively, from 55.2% to 56.8%. These gains came at the expense of a squeezed Middle 40%, whose share of income fell by the corresponding amount (3.7 percentage points) over the period. Therefore, while inequality among the bottom 90% declined, the top consolidated its concentration. Moreover, the strongest changes, in both surges and declines of income shares, came after the early 2000s, coinciding with the renewal of growth and the change of government.

Table 2.1 presents the income thresholds and averages for these income groups as well as for more refined shares at the top in 2015 USD PPP. In this year, to be one of the richest 10% of adults in Brazil one needs to make the equivalent of at least 26,977 dollars per year (almost 50,000 reais). The average income of the top decile was just over 105,000 dollars (over 196,000 reais). The magnitudes increase substantially as we move into the top percentile of the distribution, with the average income of the richest 1% being around 585,000 dollars (over 1 million reais). Table 2.2 shows the average incomes of different groups in the population in Brazil compared to those in France and the USA in purchasing power parity Euros of 2014. The structure of inequality in Brazil depicts a country of two radically different societies. While Brazil is half as affluent as France and less than one third as affluent as the USA overall, there is an extremely rich group at the top with broadly comparable levels of pre-tax income. Not only do economic elites in Brazil aspire to have similar *relative* incomes to their counterparts in rich countries, but they also manage to have the similar *absolute* levels. This is especially true with respect to their European

Table 2.1: Income Thresholds, Averages and Income Shares in Brazil: 2015

Income Groups	Number of Adults	Income Threshold (2015 USD PPP)	Average Income (2015 USD PPP)	Income Shares
Full Population	143,393,748	0	19,297	100.0%
Bottom 50%	71,696,874	0	5,256	13.6%
Middle 40%	57,357,499	10,316	15,175	31.5%
Top 10%	14,339,375	26,977	105,994	54.9%
Incl. Top 1%	*1,433,937*	*154,747*	*585,955*	*30.4%*
Incl. Top 0.1%	*143,394*	*913,201*	*3,116,227*	*16.1%*
Incl. Top 0.01%	*14,339*	*4,632,144*	*16,505,176*	*8.6%*
Incl. Top 0.001%	*1,434*	*28,509,459*	*88,697,557*	*4.6%*

Notes: This table reports statistics on the distribution of national income in Brazil in 2015. The unit is the adult individual (20-year-old and over; income of married couples is split into two). In 2015, 1 US dollar = 3.3 reals (market exchange rate) or 1.85 reals (purchasing power parity). Income corresponds to pre-tax national income. Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates combine national accounts, surveys and fiscal data.

counterparts. However, while Brazilian elites fulfil their upward-looking comparisons, the remainder of the population are far from their counterparts in advanced countries. An emblematic case is the average income of Brazil's Middle 40% being below the average income of the Bottom 50% in both France and the USA. This conveys the lack of a broad "middle class" in Brazil's dual social structure.

Table 2.2: Average incomes in Brazil, France and USA in 2014

Income Groups	Brazil (2014 Euros PPP)	France (2014 Euros PPP)	USA (2014 Euros PPP)
Full Population	**15,176**	**32,688**	**49,509**
Bottom 50%	4,131	14,692	12,422
Middle 40%	11,826	36,691	50,054
Top 10%	83,805	106,660	232,767
Incl. Top 1%	*456,908*	*352,921*	*1,000,041*
Incl. Top 0.1%	*2,481,309*	*1,208,114*	*4,614,051*
Incl. Top 0.01%	*12,431,446*	*4,226,609*	*21,550,391*
Incl. Top 0.001%	*67,443,742*	*12,894,262*	*94,063,272*

Notes: The unit is the adult individual (20-year-old and over; income of married couples is split into two). In 2014, 1 Euro = 2.32 reals (purchasing power parity). Income corresponds to pre-tax national income. Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates combine national accounts, surveys and fiscal data.

Table 2.3 presents the 2015 shares for the same income groups and across our different income series previously defined. For instance, the Top 1% (about 1.4 million adults) in the surveys received 10% of income. However, when we correct top incomes using fiscal data and factor in undistributed income from national accounts the share increases dramatically to 23% in the former and to 30% in the latter. In other words, the top percentile commands

30 times the average income of the country. The large and growing share captured by the Top 1% since the late 1990s can be contrasted to the falling shares received by both the Middle 40% share and the Top 10-1%, as Figure 2.6 shows. The latter group is commonly referred to as the country's "upper middle class".

Moving up the distribution the story is the same, with economic elites capturing disproportionate shares of total income. Figure 2.7 shows that the Top 0.1% (140,000 adults) has been receiving at least as large a share of total income as the Bottom 50% (70 million adults). Starting at similar levels in the mid-1990s both experienced similar growth until 2010, with the Bottom 50% subject to less volatility. From 2010, the Top 0.1% has surged ahead.

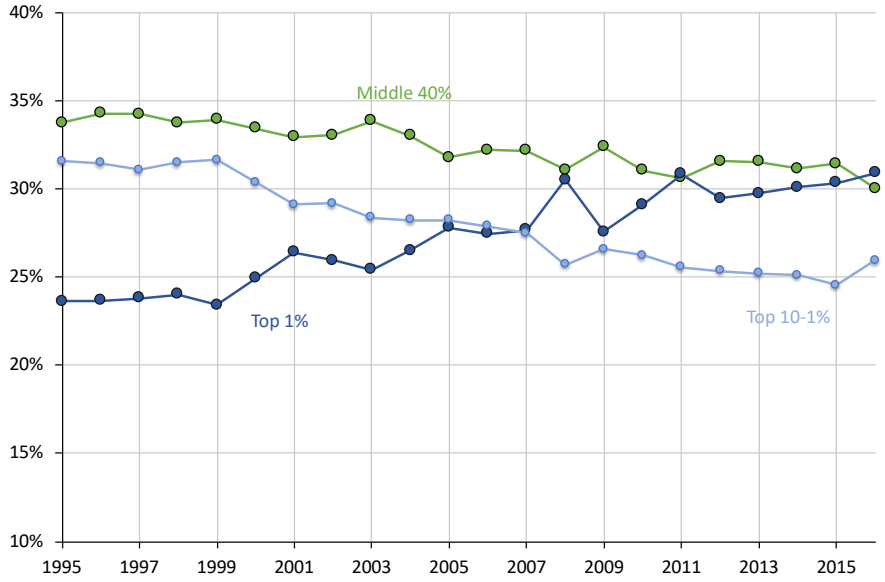
Table 2.3: Income Shares in Brazil in 2015

Income Groups	Survey income series	Fiscal income series	National income series
Bottom 50%	16.9%	13.7%	13.6%
Middle 40%	45.3%	37.0%	31.5%
Top 10%	37.7%	49.3%	54.9%
Incl. Top 1%	*10.2%*	*22.5%*	*30.4%*
Incl. Top 0.1%	*2.2%*	*10.4%*	*16.3%*
Incl. Top 0.01%	*0.5%*	*5.3%*	*8.2%*
Incl. Top 0.001%	*0.1%*	*2.6%*	*4.4%*
Total (% of national income)	55.4%	70.0%	100.0%

Notes: The unit of observation for all series is the equal-split adult. The survey income series uses the raw survey data, where the incomes match those assessed in the tax declarations . The fiscal income series combines survey and tax data. The national income series combines survey and tax data with non-fiscal income from the national accounts.

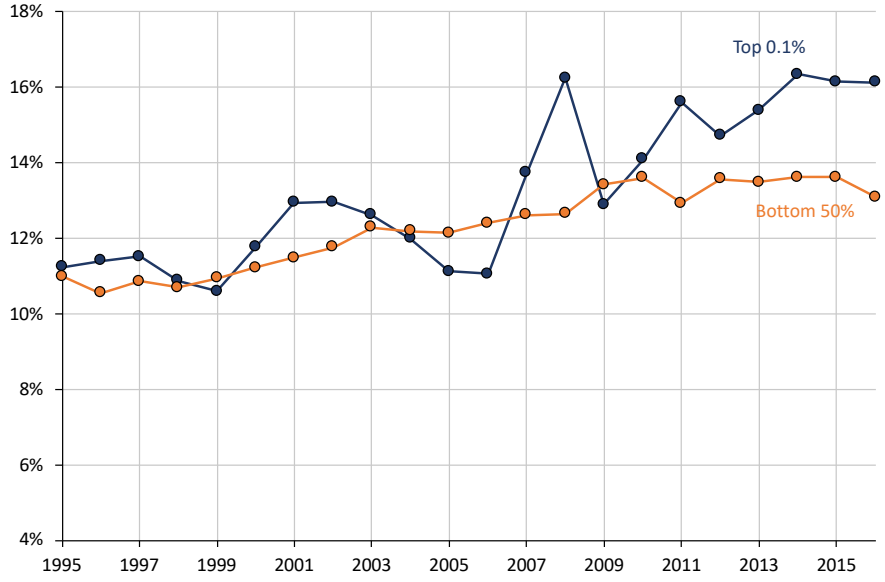
Figures 2.8–2.11 plot the temporal comparison of the estimates from our three series: the raw estimates from the surveys, our corrected series for fiscal income (combining survey and fiscal data) and our benchmark national income (DINA) series (combining national accounts, surveys and fiscal data). In all cases, what the surveys allow us to estimate is a very distorted picture of reality. When compared to our benchmark national income series, the discrepancy is very large and increasing the higher up we move in the distribution, as well as over time. In general, the bulk of the correction is made from survey income to fiscal income using information from tax declarations. Therefore, relying exclusively on surveys or even ignoring undistributed income in national accounts flowing to corporations can distort the dynamics at play. For instance, according to the surveys inequality unequivocally fell over the last 15 years (the top shares fell, while the middle and bottom shares rose). However, the national income series shows a more nuanced picture — stable concentration at the top (increasing with the deepening recession), less of an increase at the bottom and an ever-squeezed middle over the period.

Figure 2.6: Top 1% vs Middle 40% vs Top 10-1% Income Shares



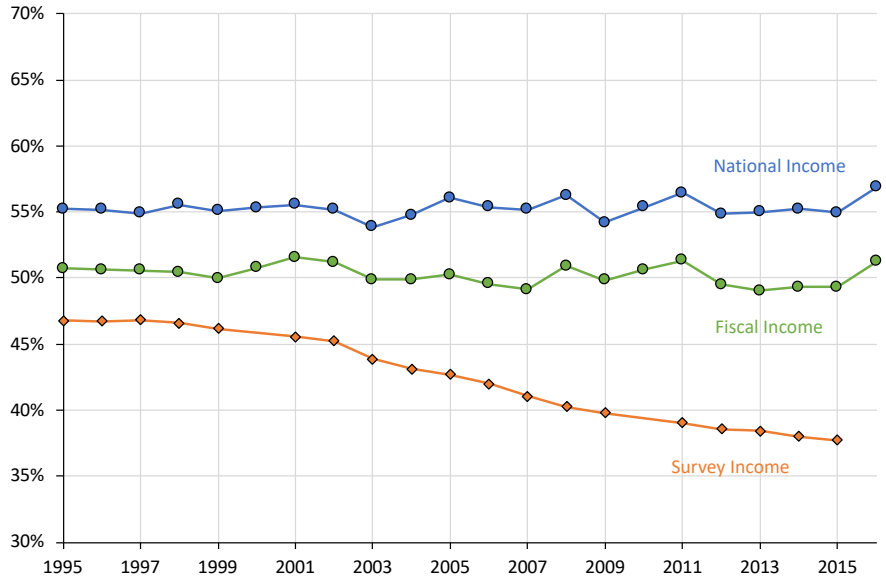
Notes: Distribution of pretax national income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining survey, fiscal and national accounts data). Equal-split-adults series (income of married couples divided by two). The Top 10-1% corresponds to the top decile excluding the Top 1%.

Figure 2.7: Top 0.1% vs Bottom 50% Income Shares



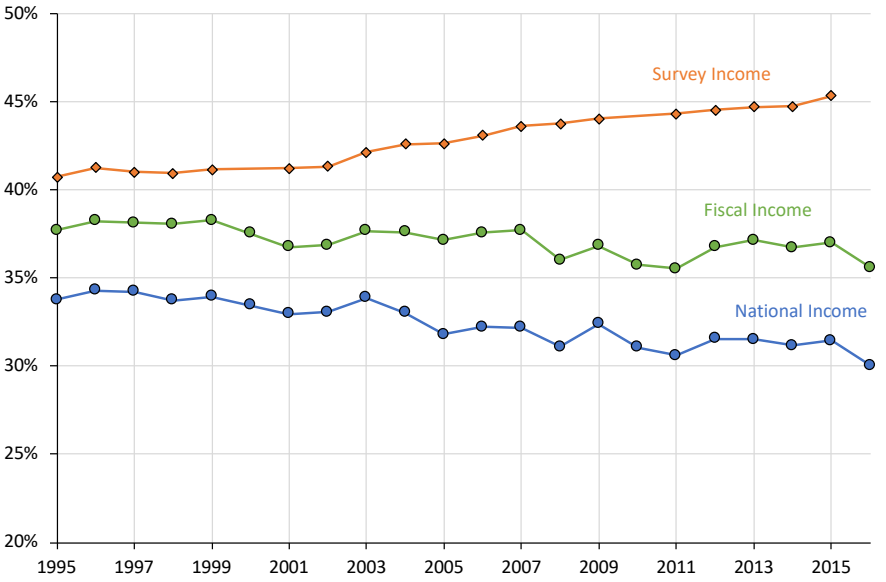
Notes: Distribution of pretax national income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining survey, fiscal and national accounts data). Equal-split-adults series (income of married couples divided by two).

Figure 2.8: Top 10% Shares Across Income Concepts



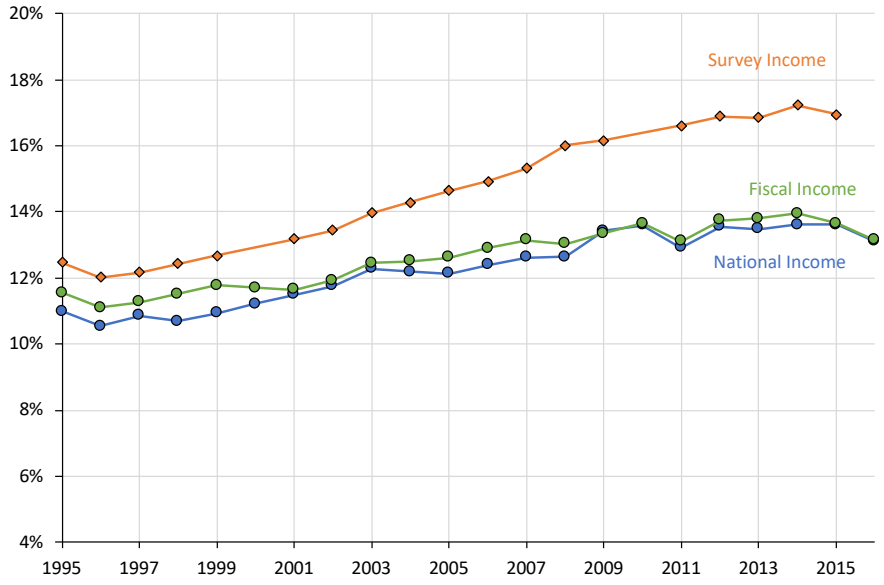
Notes: Distribution of total income (before taxes and transfers, except pensions and unemployment insurance) among adults in our three series, raw estimates from surveys, a fiscal income series (combining surveys and fiscal data) and a national income series (combining national accounts, surveys and fiscal data). Equal-split-adults series (income of married couples divided by two).

Figure 2.9: Middle 40% Shares Across Income Concepts



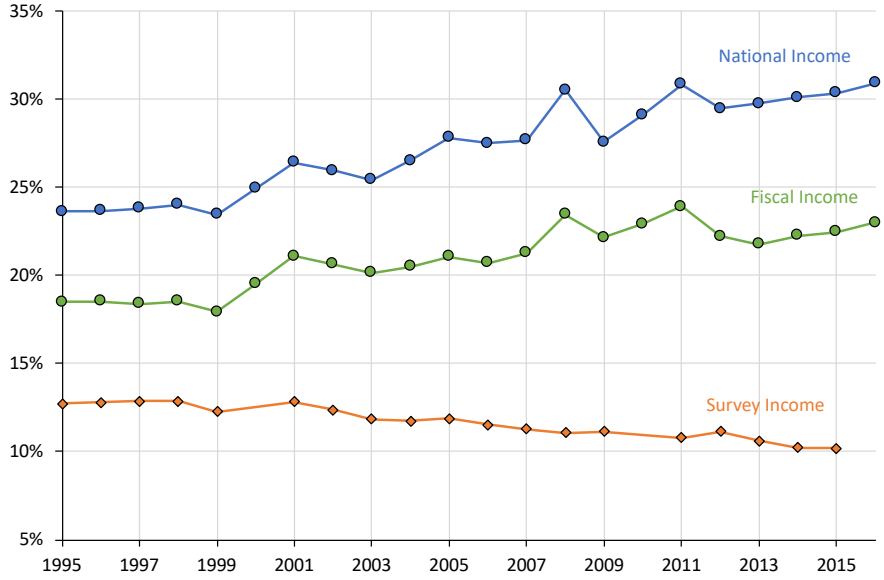
Notes: Distribution of total income (before taxes and transfers, except pensions and unemployment insurance) among adults in our three series, raw estimates from surveys, a fiscal income series (combining surveys and fiscal data) and a national income series (combining national accounts, surveys and fiscal data). Equal-split-adults series (income of married couples divided by two).

Figure 2.10: Bottom 50% Shares Across Income Concepts



Notes: Distribution of total income (before taxes and transfers, except pensions and unemployment insurance) among adults in our three series, raw estimates from surveys, a fiscal income series (combining surveys and fiscal data) and a national income series (combining national accounts, surveys and fiscal data). Equal-split-adults series (income of married couples divided by two).

Figure 2.11: Top 1% Shares Across Income Concepts

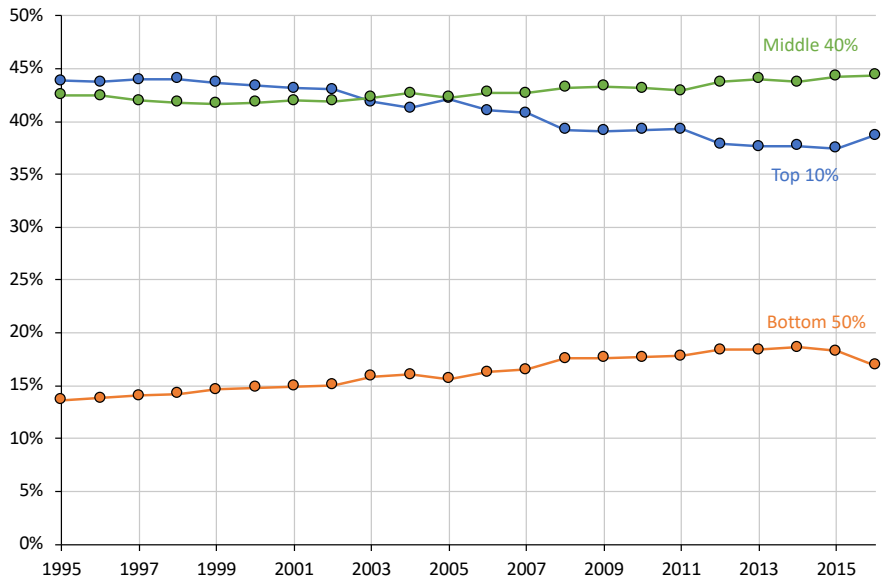


Notes: Distribution of total income (before taxes and transfers, except pensions and unemployment insurance) among adults in our three series, raw estimates from surveys, a fiscal income series (combining surveys and fiscal data) and a national income series (combining national accounts, surveys and fiscal data). Equal-split-adults series (income of married couples divided by two).

3.2 The Decline that Was: Labour Income Inequality

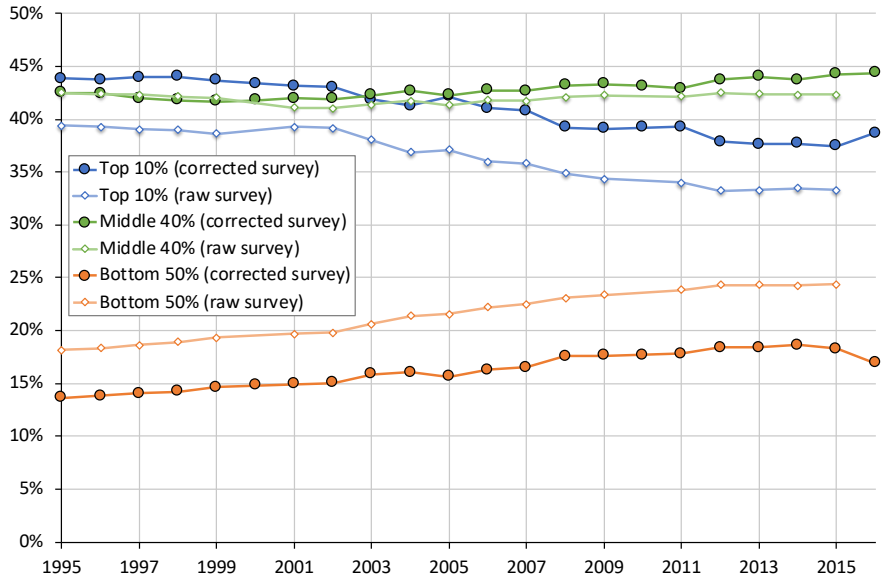
Despite the overall distribution of income failing to narrow as much as "official" sources pertained, there is evidence that the inequality of labour earnings declined, even when using administrative data to correct surveys. Figure 2.12 presents our shares of labour income when using income tax declarations to correct the top of the labour income distribution from the surveys (as described in section 2.8). The Top 10% labour income share fell steadily (by 12%), the Middle 40% share increased (by 4% since 1995, and 6% since 2002), while the Bottom 50% labour income share increased considerably more than the national income series (24% vs 19%). Again, surveys seem to underestimate the dispersion in labour income (Figure 2.13), although by a lower magnitude than for total income (compare with Figures 2.8–2.10). The decline in "wage inequality" is more apparent if we focus on formal employees (excluding civil servants) using data from social contributions. The Top 10% wage income share falls more strongly over the period than the share of total labour income (which includes labour benefits, wages of informal and self-employed workers and public servant salaries), while the Bottom 50% share increased twice as much (see Figure 2.14). The evolution is consistent with that observed using the raw survey data (Figure 2.15), but the levels of inequality are still higher in the administrative data. The clear picture that emerges from this analysis is that the squeezed group is now the top, rather than the middle, and that the evolution of formal salaries contributed a substantial positive element to the decline in labour income inequality, which was strongest from 2002 onwards.

Figure 2.12: Labour Income Shares in Brazil: Corrected Estimates



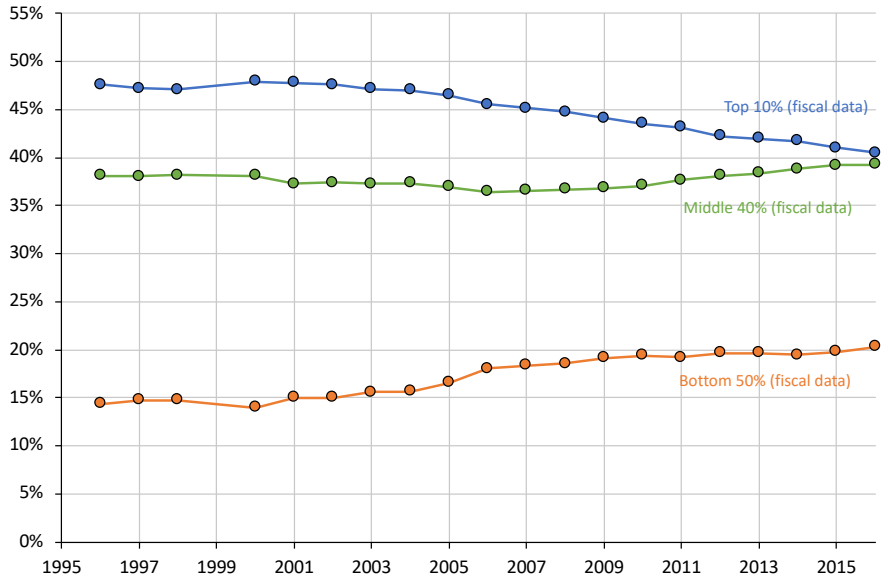
Notes: Distribution of pre-tax labour income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining surveys and tax data). Equal-split-adults series (income of married couples divided by two).

Figure 2.13: Labour income shares in Brazil: Corrected vs Raw Estimates



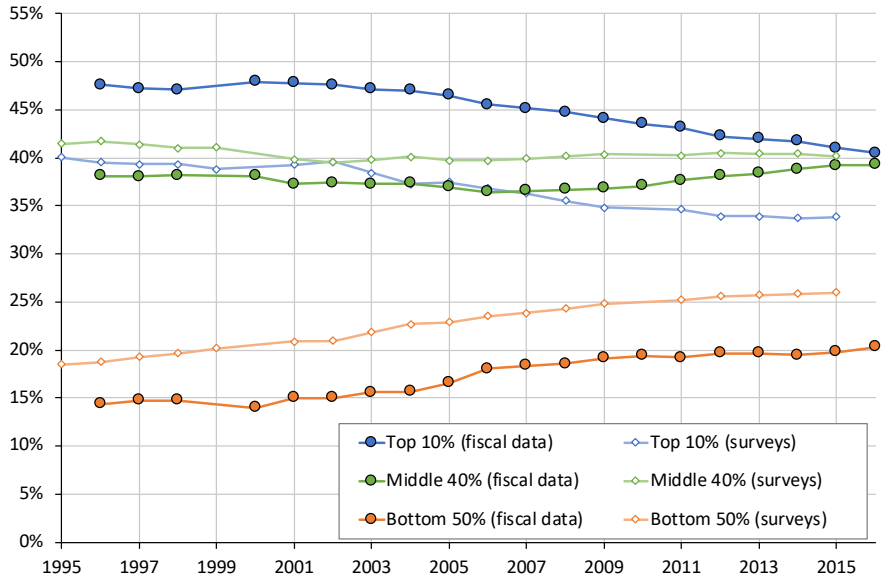
Notes: Distribution of pre-tax labour income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining surveys and tax data). Equal-split-adults series (income of married couples divided by two).

Figure 2.14: Shares of Formal Sector Earnings in Brazil



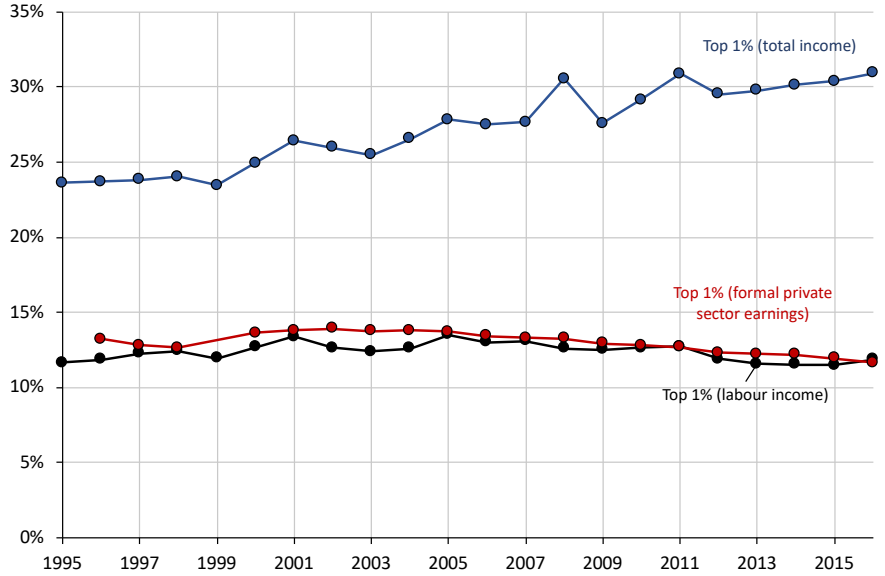
Notes: Distribution of wages and salaries of all formal sector employees appearing in social contributions data. These account for about 40% of the adult population.

Figure 2.15: Shares of Formal Sector Earnings: Surveys vs Fiscal Data



Notes: Distribution of wages and salaries of all formal sector employees from social contributions data compared with the distribution observed in the surveys.

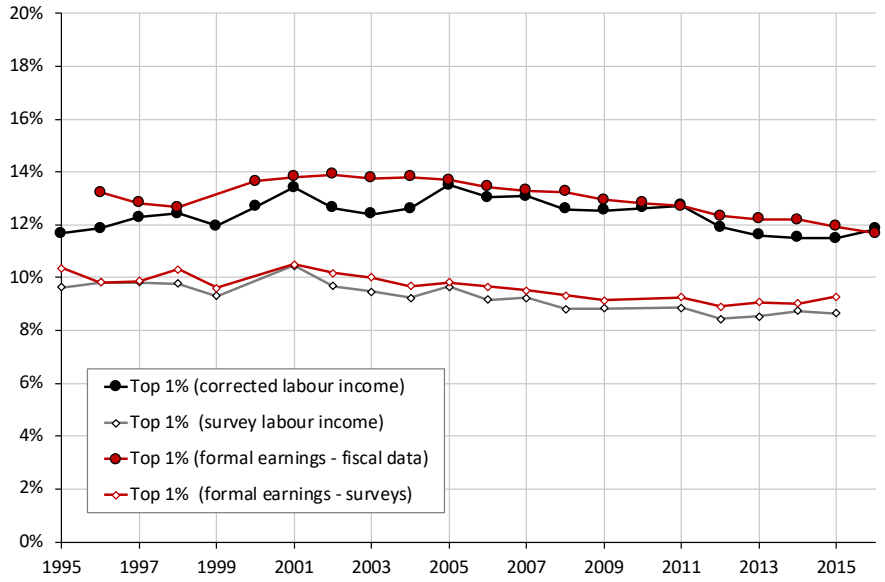
Figure 2.16: Top 1% in Brazil: Total Income vs Labour Income vs Earnings



Notes: Top 1% of pre-tax national income (before taxes and transfers, except pensions and unemployment insurance) among adults compared to the corrected labour income series (combining surveys and tax data) and to the formal earnings series (from social contributions data). Equal-split-adults series (income of married couples divided by two) for the national income series and labour income series and individual employees for formal earnings series.

Figure 2.16 conveys the evolution of labour income and earnings among the Top 1%, compared to our benchmark total income series. First, the levels of concentration at the top percentile of our two labour income distributions is much less than the equivalent concentration in the total income distribution. Second, their evolution followed a visible downward trend only from 2005, which is broadly consistent with what the surveys show but along different levels (see Figure 2.17). These dynamics imply that the distribution of the returns to capital ownership played an important role in limiting the fall in total pre-tax inequality in Brazil.

Figure 2.17: Top 1% Labour Income in Brazil: Corrected vs Raw Estimates

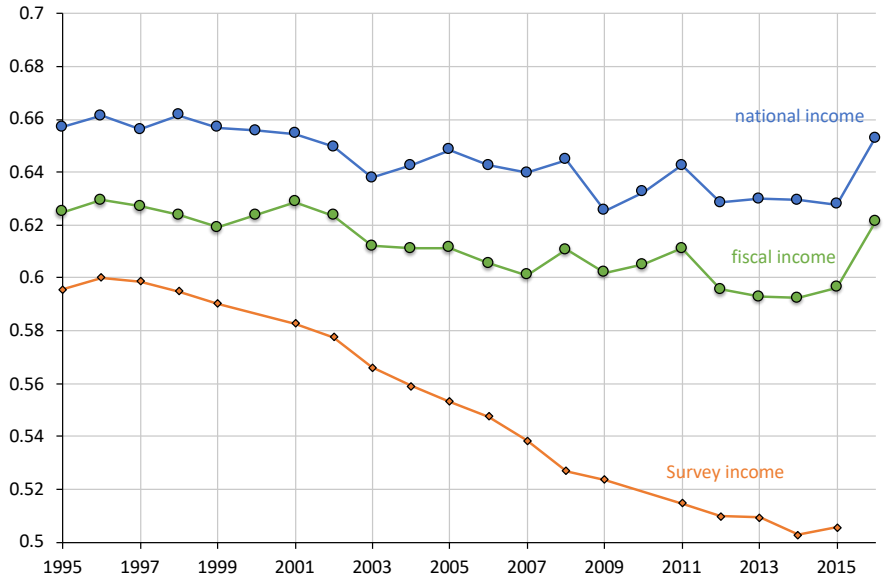


Notes: Comparison of top 1% corrected labour income (combining surveys and tax data) and formal earning series (from social contributions data) to raw series from household surveys. Equal-split-adults series (income of married couples divided by two) for the labour income series and individual employees for formal earnings series.

These findings are consistent with the evolution of the Gini coefficients, as depicted in Figure 2.18. Overall inequality measured in this way registered a decline, but does so most strongly for the survey income distribution, which falls by almost 10 percentage points (p.p). Given that surveys are more reliable in reporting labour incomes, which experienced a strong compression, this is unsurprising. Between 2001 and 2015 the Gini of national income and fiscal income both decreased by about 3 p.p. The deepening of the recession in 2016 seems to have cancelled the gains made over the previous fifteen years.

However, it must be noted that the Gini measure of inequality is less sensitive to movements in the extreme tails of the distribution. This implies that it mostly captures changes in the distribution of labour earnings, given the deduced concentration of capital income in the right tail. Figure 2.19 more clearly illustrates this – Gini trends for corrected

Figure 2.18: Gini coefficients for Total Income in Brazil: 1995–2016



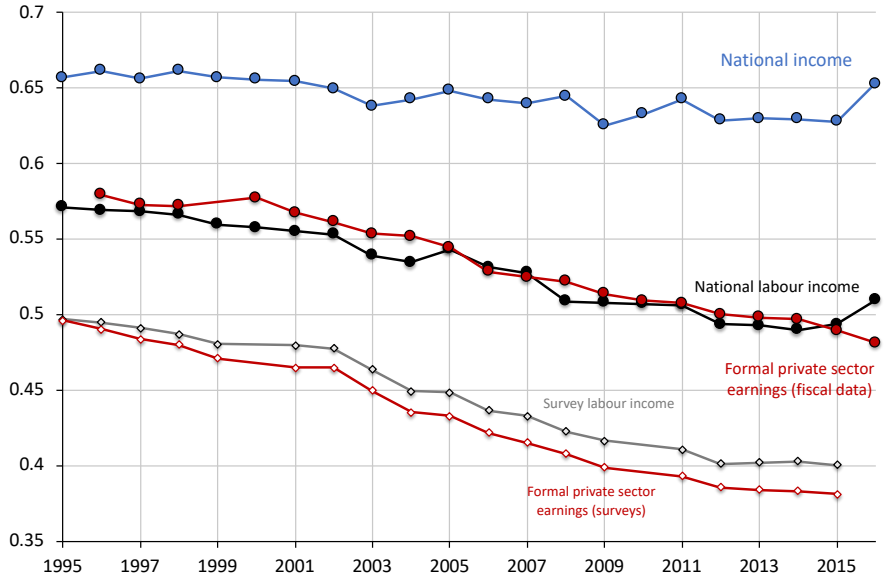
Notes: Distribution of pre-tax income among equal-split adults across income concepts. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data). Survey income is computed from raw survey data. Fiscal income combines surveys and tax data. National income combines fiscal income from surveys and tax data with non-fiscal income from national accounts.

labour income and formal earnings from fiscal data closely follow their survey-based counterparts. Therefore, it is not wholly unsurprising that the corrected national income Gini coefficient shows a slight decline over time, given that labour income is its principal component (accounting for about two-thirds of total income, and closer to 100% of income for 90% of the population). Section 4 discusses how we can understand these evolutions in the political context of the period.

Despite the fact that we are mainly interested here in the evolution of primary income inequality, before the intervention of government transfers (except social security pensions and unemployment insurance), we can perform a simple calculation to factor into our estimates the extent to which the country's celebrated targeted cash transfers make a difference. Brazil's cash transfers have received much media attention for their positive redistributive impact on the poor. Indeed, the published research on Brazil to date has placed much emphasis on the increased resources that have been dedicated to social assistance programs since the early 2000s (Barros et al., 2010).

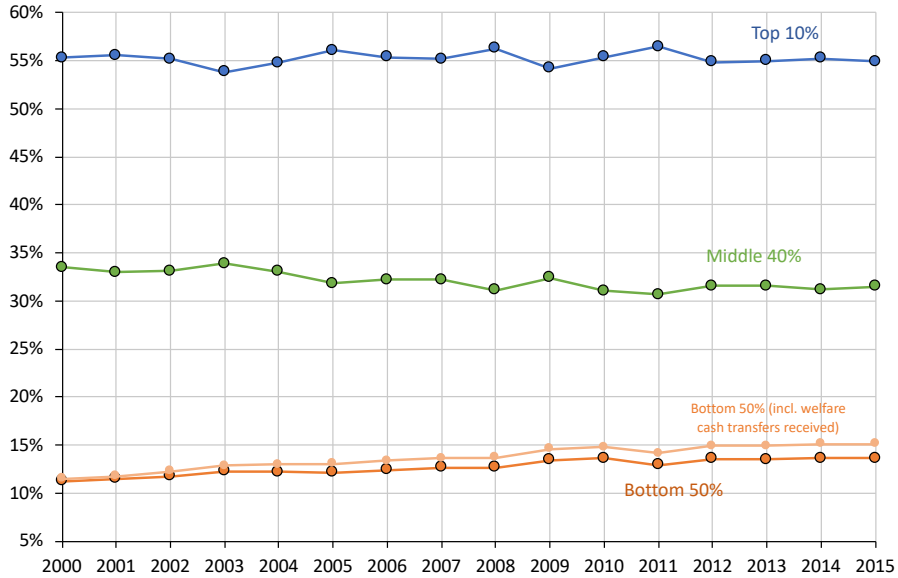
To better see their impact, we make the following calculation. From national accounts, we obtain the annual amount of social assistance cash transfers received by households. These include Brazil's flagship conditional cash transfer, Bolsa Familia, and welfare pensions for the elderly or incapacitated poor (*Benefício de Prestação Continuada*). These programs

Figure 2.19: Gini coefficients in Brazil: Total Income vs Labour Income



Notes: Distribution of pre-tax income among equal-split adults across income concepts. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data). National labour income includes all proceeds from salaried and independent work, and labour benefits (pensions, unemployment insurance, annual bonuses and other top-ups).

Figure 2.20: The Impact of Cash Transfers in the Brazilian Distribution



Distribution of pretax national income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining survey, fiscal and national accounts data). Equal-split-adults series (income of married couples divided by two). The annual share of social assistance transfers in national income is added to the share of the Bottom 50%.

accounted for about 1% of national income, varying from 0.3% in 2000 to 1.5% by the end of the period. Given the difficulties in estimating the exact incidence of these transfers (due to them being financed out of general tax revenues) we show the impact of these transfers for the broad income group that were the gross recipients of them – the Bottom 50%. The results of this calculation are presented in Figure 2.20. The share of the Bottom 50% increases from 11.5% in 2000 to 15.1% in 2015. The stronger growth in the share of the poorest 50% including welfare transfers is due to governments dedicated greater resources to them over the decade and half. It can be seen that what matters for the income share of the poor is the size of these programs in national income. Although they make a big difference at the household level, in the aggregate distribution their contribution is still slim due to their marginal share in the economy. The actual "net impact" of these transfers is evidently more complex to estimate. This is because the government increased its public spending largely out of growing public revenues (from existing taxes and public enterprises), as the economy and (formal) employment expanded (Gobetti and Orair, 2015). Thus, the fungible nature of these welfare transfers means we cannot identify a specific source of funding.

3.3 The Distribution of Growth and Recession

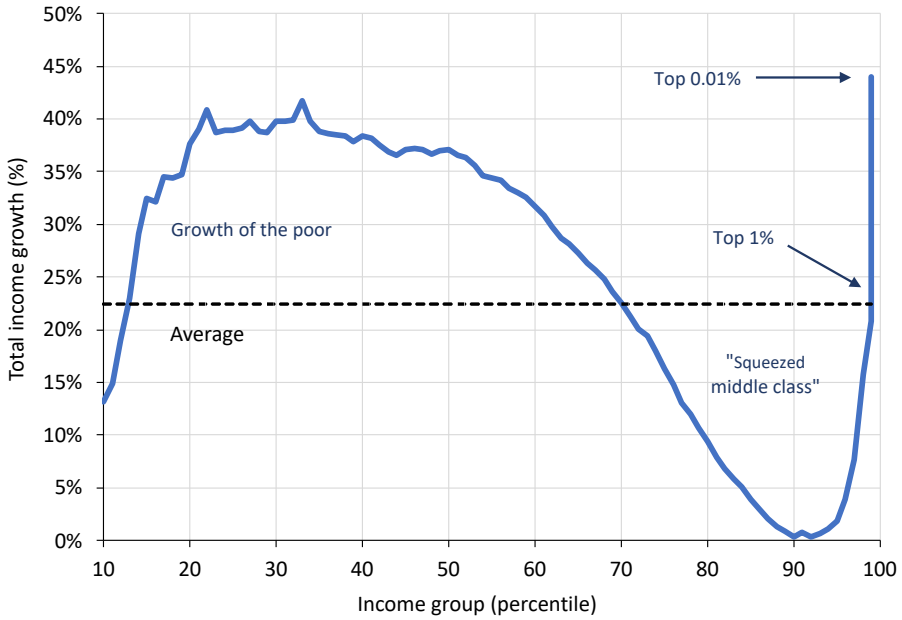
Table 2.4: Income growth, recession and inequality in Brazil: 2002-2016

Income Groups	Total Cumulated Growth (2002–2016)	Total Cumulated Growth (2002–2013)	Total Cumulated Growth (2013–2016)
Full Population	**4.7%**	**21.0%**	**-13.5%**
Bottom 50%	16.6%	38.8%	-16.0%
Middle 40%	-4.9%	15.4%	-17.6%
Top 10%	7.9%	20.6%	-10.5%
Incl. Top 1%	*24.7%*	*38.7%*	*-10.1%*
Incl. Top 0.1%	*30.4%*	*43.8%*	*-9.3%*
Incl. Top 0.01%	*38.8%*	*43.9%*	*-3.6%*
Incl. Top 0.001%	*83.1%*	*82.5%*	*0.3%*

Notes: Distribution of pre-tax national income among equal-split adults. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal and national accounts data).

As evidenced in the previous sections, the strongest changes in the distribution came after 2002. Between 2002 and 2013, average income grew by 21%, after which the incomes per adult receding by more than half of the previous decade's gains. The question that arises from this evolution is how the average income growth of different income groups fared in the two distinct periods. Table 2.4 presents my calculations for the same broad income groups treated before. Consistent with evolution of income shares, the average income

Figure 2.21: Growth Incidence Curve in Brazil: 2002–2013



Notes: Distribution of pretax national income (before taxes and transfers, except pensions and unemployment insurance) among adults. Corrected estimates (combining survey, fiscal and national accounts data). Equal-split-adults series (income of married couples divided by two).

growth rate of the Bottom 50% was strong, compared to the Middle 40% and the Top 10% since 2002. The Middle 40% was the only group to grow at a rate less than the average for the whole population during the whole period. Growth was also strong among the top percentiles with the income of the Top 1% growing by 38.7% during the high growth decade, the same as the average growth registered by the poorest 50% over the period. The tail-symmetric nature of Brazil's growth incidence during its "mini miracle" period is more clearly illustrated in Figure 2.21. Over the course of the three PT-led executives, and expenditure-driven expansion, the growth of the poorest percentiles was surging, as was that of the richest percentiles, above the Top 1%. The graph reveals most emphatically the "squeezed middle class" – the portion of the distribution approximately between percentile P70 and P99 in the distribution. The owners of capital benefited the most from the buoyant economic conditions, with the very summit concentrating the highest growth rates.

The final column of Table 2.4 shows how the income effects of the deepest domestic recession since the early 1980s were distributed. Average national income fell by 13.5% between 2013 and 2016, but the decline was strongest for the Bottom 50% and Middle 40%, with their average incomes falling by more than the country average. On the other hand, the average income of top groups fell by less than the national average, but still declined in real terms. This explains the rising inequality trends during these years. According to my estimates, the richest 1,500 Brazilians in the Top 0.001% even made marginal gains. This confirms the view that domestic recessions have a stronger proportional impact on

the poor than the affluent, at least over the short run. This is likely because the rich have more diversified income channels and more control over their final remuneration, as well as the remuneration of their "employees".

Table 2.5: Actual Sharing vs Equal Sharing of Growth in Brazil: 2002-2013

Income groups	Actual growth	Fraction of total growth captured (actual sharing)	Counter-factual growth	Fraction of total growth captured (equal sharing)
Full Population	**21.0%**	**100%**	**21.0%**	**100%**
Bottom 50%	38.8%	16.6%	161.4%	50%
Middle 40%	15.4%	28.8%	30.1%	40%
Top 10%	20.6%	54.6%	-14.3%	10%
Incl. Top 1%	*38.7%*	*36.6%*	*-20.5%*	*1%*
Incl. Top 0.1%	*43.8%*	*19.8%*	*-21.8%*	*0.1%*
Incl. Top 0.01%	*43.9%*	*9.8%*	*-22.1%*	*0.01%*
Incl. Top 0.001%	*82.5%*	*6.3%*	*-22.1%*	*0.001%*

Notes: Distribution of pre-tax national income among equal-split adults. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal and national accounts data). Counter-factual growth rates assume that each income group captures the fraction of growth that corresponds to its population size.

Despite the gains made by the bottom of the distribution, the top captured a disproportionate part of the income growth over the expansionary period, with the Top 10% capturing 55% of total average growth and the Top 1% capturing more than half of this (Table 2.5). The bottom line is that even with the strongest growth performance (out of the broad income groups) over the period, the Bottom 50% did not capture most of the growth due to their extremely low levels of income and their subsequently low share of income. Thus, over a short-to-medium run, the income growth of the poor seems to matter less than their share of total income. This is partly why the 1.4 million richest adults in Brazil captured a higher fraction of total growth than the poorest 70 million Brazilians. For the Middle 40% it is their weak growth performance that makes the difference.

In Table 2.5, I also show the growth rates that would have been needed in order for all the income groups to have captured an equal share of total per adult growth. The counterfactual scenario shows that a substantial transfer of per adult growth from the Top 10% to the lower fractiles would have been needed for this equal sharing to occur. The Top 10% would have needed to grow by -14% (instead of 21%), the Middle 40% by +30% and Bottom 50% by 161% (instead of 15% and 39% respectively). This would have evidently needed policies targeting greater pre-tax income growth for the bottom 90%, such as more and better-paid formal jobs, as well as more regulated income growth for the top, coming (for instance) from stricter collective bargaining arrangements in firms and more binding personal income taxes. But then, some may ask: would the economy still have expanded by 21%? An closer assessment of the growth process of the 2000s and policies behind it may help to answer this question.

4 Contextualising Recent Trends in Inequality

To make sense of the evolution of inequality in Brazil in the 2000s we must comprehend the economic and political contexts of this period. Between 2000 and 2015, according to the IBGE's system of national accounts, Brazil's economy expanded by 43%, with 83% of this growth accounted for by consumption spending (private consumption making up 63%) and 17% coming from investment. Despite favourable terms of trade, net foreign demand contributed close to 0%.[31] On the production side growth was led mainly by commodity refinement, construction (in real estate and infrastructure), and the expansion of employment in services.

Since the late 1990s, in the context of the millennium development goals, poverty reduction became central to the agenda of governments in Brazil. With the election of the PT, this was given greater impetus. Their general discourse, which was mirrored in their policies, focused largely on the bottom of the distribution. Without modifying the ownership of capital in the economy, or reforming the tax system — in which company profit withdrawals, dividends and financial incomes are exempt from the progressive income tax schedule; inherited fortunes across states are taxed at an average rate of about 4%; and greater fiscal burden is placed on consumers of basic goods and services — the policy focus of the PT centred around redistributing the proceeds of production through cash transfers and increasing the bargaining power of workers through unions and collective wage negotiations, anchored to a growing real minimum wage.

While the PT governments focused much attention on the bottom of the distribution, without infringing on the privileges of economic elites, the evolution of the Middle 40% income share may seem of residual importance, particularly as the share of the Middle 40% in labour income made gains during the same time. But the importance of the Middle 40% in a developing country like Brazil should not be overlooked. By capturing little or no part of the capital income distribution, and by not capturing much of the fruits of social policy directly, "the squeezed middle" is an important political phenomenon, as the 2018 Election made apparent (Gethin and Morgan, 2018). This is particularly the case when elites succeed in placing the middle in competition for "scarce" resources with the bottom. How this middle group perceives its relative stagnation can be the key to determine the formation of political coalitions around the issues of growth and income distribution.

Furthermore, our results for the middle of the distribution challenge the hypotheses (at least for Brazil) that the middle and the tails roughly split income 50-50 (Palma, 2011, 2016), and that given the greater capacity of the middle to maintain their half, it is the bottom of the distribution that "needs to be squeezed accordingly" (Palma (2016), p. 61). Rather, it seems that the 50-50 rule applies between the top decile and the rest of the

[31] Even if the contribution of net exports to growth was nil, favourable terms of trade could have aided the importation of capital and intermediate inputs, which could have favoured investment and consumption.

population, with the balance tipped in favour of the top in Brazil. What these dynamics do point towards (consistent with Palma (2011)) is the crucial role of the income share of the top of the distribution, how it is determined and what they (or the government) decide to do with it.

4.1 The Progressive Role of Social Policy

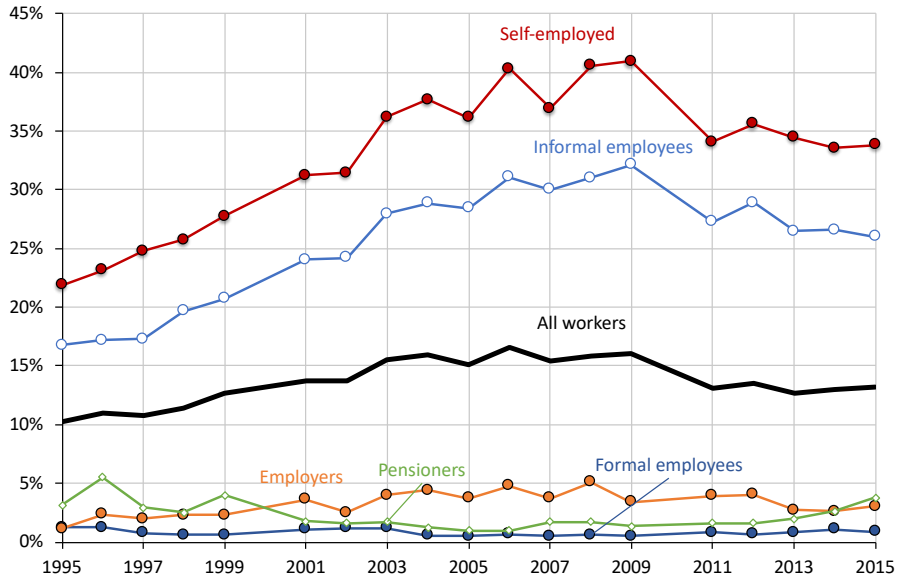
Given the methodological framework employed in this chapter, I dwell on policies affecting "market income" distribution. Here the role of remuneration practices is important. This is particularly true regarding legislation on the level of the minimum wage. In a context like Brazil's, the level and enforcement of the minimum wage becomes a crucial policy area to determine the share of income appropriated by the Bottom 50% of the population.[32] From the mid-1990s the minimum wage began to finally rise in real terms. Between 1995 and 2002, the real minimum wage increased by 2% per year. An more explicit revalorisation of the minimum wage was implemented during the two Lula governments. Between 2002 and 2010 the minimum wage increased by 4% per year in real terms (from a relatively low level of BRL 195 per month, or 227 USD PPP), while average adult income grew by 2% per year (and the median income grew by approximately 2.5% per year). With the incidence of the minimum wage being concentrated in the Bottom 50% of the income distribution, its income share was pulled upwards over the period (Ferreira et al., 2017; Brito et al., 2016). From 2010 to 2016, despite the slowdown in the economy, the minimum wage continued to grow at above average and median rates, (+1.5% per year vs. -1.8% and -0.7% per year respectively). The indexation of the minimum wage to all labour-related benefits, particularly to pensions, also acted as a progressive lever (Brito et al., 2016).

Enforcement of the minimum wage also seems to have improved in recent years, with enforcement levels varying widely among different categories of workers. According to information from the PNAD, we can estimate that less than 15% of total workers earned less than the minimum wage on average in their principal occupation since 1995, with a decline observable since 2006 (see Figure 2.22). Before the mid-2000s, there is evidence of declining compliance. Self-employed workers and informal employees were the most prone to earn below the minimum wage (33% and 26% earning less respectively), while employers, pensioners and formal employees were the least likely to be earning less than the minimum legislated threshold (all less than 5%). The fact that the majority of informal employees (which represent over 40% of total employees) and self-employed workers are covered by the minimum wage, shows the spillover effects that it has across the workforce.

Concerning the documented reduction in labour income inequality (see section 3.2), various other factors can be pinpointed. The first is the reduction of the education premium, which has been linked to the increase in the relative supply of skilled workers, completing

[32]The average income of the Bottom 50% in Brazil has closely matched the level of the minimum wage since at least the 1970s (see chapter 3).

Figure 2.22: Share of Workers Earning Less than the Minimum Wage



Notes: workers with positive earnings in principal occupation from PNAD. Employees include domestic workers. In 2001 1% of formal employees earned less than the minimum wage in their principal occupation, while among all workers this rate was 14%. In the same year, the share of employees (including domestic workers) without a formal contract was 47%.

secondary or tertiary education, (Barros et al., 2010), and represents the first significant reduction in Brazil's history. This was on the back of important educational targets in the 1988 Constitution, which committed defined spending thresholds by different tiers of governments. Compared to other countries across different 20-year time periods, Brazil underwent one of the largest educational expansions ever recorded from 1990 to 2010 (Barro and Lee, 2013). Since 1990, enrolment rates in primary education increased from 85% to practically 100%, while for secondary education it increased remarkably from 16% to 86%. The enrolment rate in tertiary education increased from about 15% to 30%. The 2000s were met by both an increase and equalisation in the funding of basic education (primary and lower secondary schooling) across all regions and the creation of a program that subsidised the education for children from poor backgrounds, initiated as *Bolsa Escola* under the Cardoso administration in the 1990s, but later integrated by the first Lula government into the broader *Bolsa Familia* program (Bruns et al., 2011). At the same time, the falling returns to education were also influenced by the increase in the relative demand for unskilled labour associated to certain growing economic sectors, such as construction and services (Lustig et al., 2016). Thus, contrary to rich country experience, unskilled-bias technical change was important in Brazil. We can add to the educational mix the increase in resources allocated to active labour market programs and retraining since the 2000s. By 2015 the PT had doubled the share of spending in these areas.

Other documented factors linked to the fall in labour earnings inequality in the PNAD

are related to the decline in "horizontal wage inequalities". These consist of the inequality between men and women; blacks and whites; rural and urban areas; and formal and informal sector workers, conditional on the evolution of the minimum wage and educational attainment (Ferreira et al., 2017). Informality rates in general fell from 40% of employed workers in the early 2000s to less then 30% by 2015. Not only were less privileged workers gaining greater labour market protection, but successive PT governments improved the monetary conditions of low-wage workers through labour benefits such as the *abono salarial* (a bonus of 1/3 of a minimum wage for employees making less than 2 minimum wages) and the 13^{th} monthly salary and pensions, all linked to the policy of minimum wage revalorisation. Ferreira et al. (2017) argue that falling returns to *experience* was in fact a more important factor in the decline of labour earnings, which goes in line with the unskilled-biased evolution of the 2000s. Overall, the evolution of minimum wages and indexed labour benefits made returns to education and experience less important.

Given the variables driving Brazil's growth surge up to 2014, it is not evident that a more equal sharing of growth among the population would have necessarily compromised the trajectory. With growth coming from domestic expenditures (mainly consumption, but also investment), stronger income growth for the Bottom 90% in the distribution could have created greater aggregate demand effects on overall growth, under the (reasonable) assumption that the marginal propensity to consume of lower-income groups is higher than that for higher-income groups.[33] However, given the distributional conflict and political polarisation that has emerged in recent years, achieving a decline in the average income of individuals within the top decile and higher fractiles of 14-22% over a decade (as in Table 2.5), could have produced an even more delicate political economy counter-factual. In hindsight, the political viability of such a distributional shift would have depended on a broadly united voting coalition among the Bottom 90% of the distribution.

4.2 The Regressive Role of Taxation and Monetary Policy

As alluded to, progressive policy largely bypassed the domain of taxation. The highly stratified income tax system was inherited without any significant reforms by the governments of the 2000s. Personal capital and labour incomes continue to receive unequal treatments in the tax code. This form of inequality could have had non-negligible effects for the distribution of pre-tax income. Not only does it incentivise aggressive bargaining at the top for higher pay (Piketty and Zucman, 2014), but changes in the *form* of this pay also matters for the assessment of inequality trends.

[33]Data from the IBGE's 2009 Survey of Family Budgets (the *POF* survey) confirms the accuracy of this assumption, with consumption representing a larger share of income for poorer individuals. https://www.ibge.gov.br/estatisticas-novoportal/sociais/saude/9050-pesquisa-de-orcamentos-familiares.html. For those earning more than 22 minimum wages per month (approximately the Top 1%) consumption represents about 56% of their gross income, while it represents more than 100% for those earning less than 1.8 minimum wages per month (approximately the Bottom 40%) ((Carvalho, 2018), combined with own calculations for fractiles).

While it is clear that labour income inequality fell over the period, it is not possible to fully disentangle the causes of the change. One the one hand, the factors outlined above, like the minimum wage, education and between-group inequalities, were present and are relevant for the analysis. But it is also possible that the decline in wage inequality could be explained, in part, by a change in remuneration forms among individuals with a high degree of control over their pay (business owners, including self-employed professionals and corporate executives). If these workers changed their remuneration away from wages to capital withdrawals like dividends, then labour inequality could have fallen if this behaviour was concentrated at the top of the distribution. In a context where distributed profits are exempt from the personal income tax (while top labour incomes are taxed at the top marginal tax rate of 27.5%) this type of behaviour should not be ruled out, especially as it seems common among business-owners and those in self-employment.[34] This may help to explain why top shares of total fiscal income have been stable despite the fact that the shares in its main component – labour income (representing at least 75% of total fiscal income) — have declined.

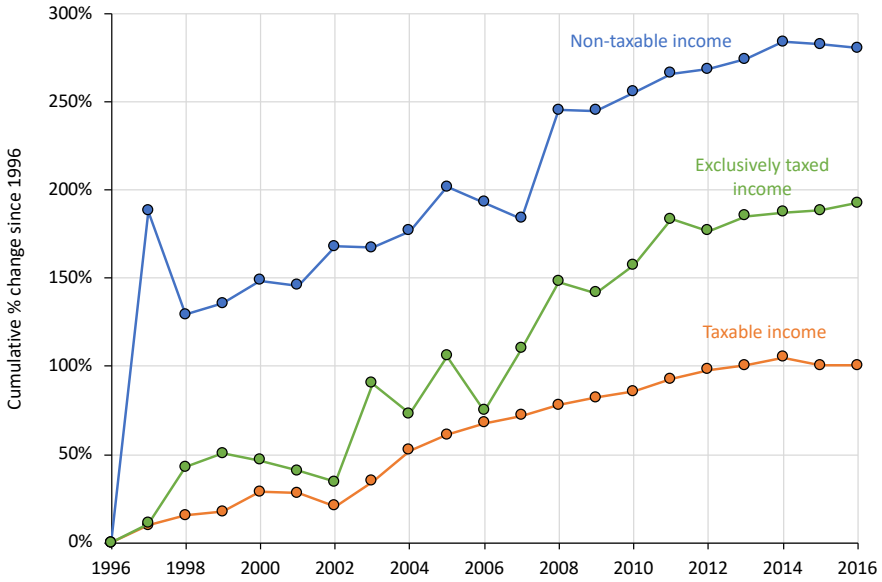
Evidence of these potential income shifts can be drawn from income tax statistics. Figure 2.23 shows that since 1996, the year from which profits and dividends were exempted, nontaxable incomes have vastly outgrown taxable incomes. Moreover, between 2007 and 2016, it is possible to directly estimate the growth of aggregate profits and dividends distributed to partners and shareholders from the tax returns (see Figure 2.24). These incomes grew more than twice as much as wages and salaries. At the same time, total declarations grew by 1% per year, while the number of people declaring dividends and profits grew by 10% per year.[35] These dynamics resulted in the share of wages/salaries in total income to fall from 70% in 2007 to 59% by 2016, while the share of profits/dividends rose from 11% to 13%.

Moreover, this transfer seems to have been concentrated across upper-middle class groups (independent workers registering themselves as micro-enterprises) and economic elites (corporate executives and shareholders). Figure 2.25 not only shows that non-taxable incomes are more important as one moves up the total income distribution, but the difference with taxable income has increased between 2007 and 2015. Thus, affluent groups seem to have shifted at least part of their income from taxable to non-taxable sources over time. This is suggestive evidence that the fall in labour income inequality during the 2000s could, in part, have been due to behavioural changes of affluent individuals, motivated by tax incentives in a context of rising labour costs.

[34]Distributed business profits and dividends have been exempted from contributing to the personal income tax since 1996 (Lei no. 9.249/1995). Previously they were taxed at source at a flat rate of 15%. See Nóbrega (2014). On the basis of commonly circulated anecdotal evidence, most own-account workers in Brazil with sufficient means operate as legal business and register their income as profit withdrawals or dividends to minimise their income tax payments.

[35]Total declared income grew by 5% per year, wages and salaries by 3% per year and profits and dividends by 7%.

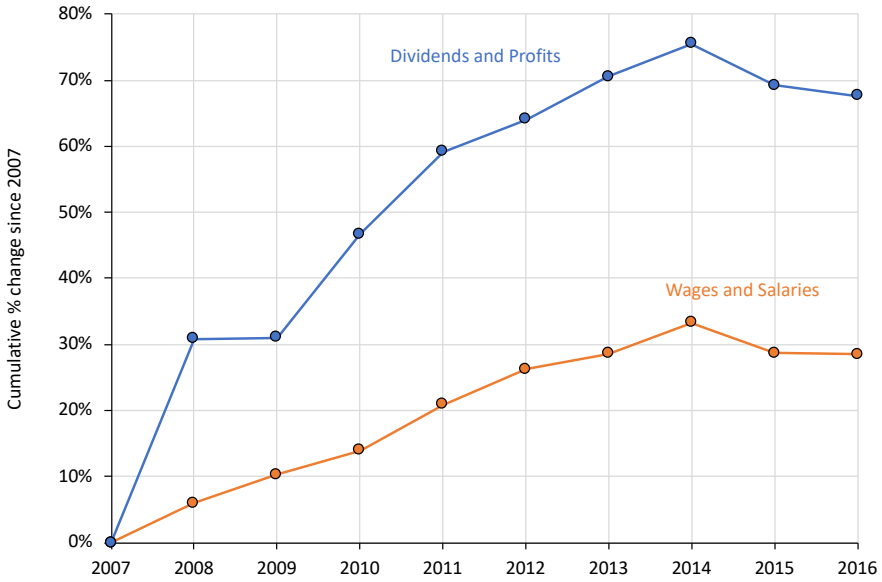
Figure 2.23: Evolution of Fiscal Income Categories in Tax Declarations: 1996–2016



Notes: Cumulative growth of the three broad income categories reported in tax data. "Taxable income" is mainly composed of the wages/salaries of employees and self-employed workers and the taxable portion of pensions, as well as property rent. It is subject to the personal income tax schedule (with marginal rates from 0-27.5%) after deductions. "Exclusively-taxed income" mainly consists of the 13^{th} salary and financial incomes (interest, capital gains), taxed exclusively at source at flat rates in the case of financial incomes and the personal income tax schedule in the case of the 13^{th}. "Non-taxable income" has distributed profits and dividends as its largest component, as well as the capital withdrawals of micro-enterprises, alongside other labour and capital receipts. See section 2.3.

While the aforementioned factors are important in explaining the evolution of fiscal income inequality, the evolution of national income inequality takes into account other components, whose distributions should not be ignored. Primary among non-fiscal capital incomes are the undistributed profits of corporations belonging to households, which represent about 6% of national income and are highly concentrated in the Top 1%, based on what we can deduce from distributed profits and the distribution of financial incomes from our data sources. Between 2000 and 2015, undistributed corporate profits grew at three times the rate of total employee compensation and total net profits (9% per year vs. 3% per year for the latter two, respectively). This increase in corporate savings occurred in a context of relatively strong macro growth amid periods of high business uncertainty (the Latin American debt and currency disruptions of the early 2000s, the global financial crisis of 2008 and the domestic recession and political tension after the 2014 election). In effect, corporate owners preferred to accumulate wealth inside their corporations as opposed to distributing themselves profits in the form of dividends (which increased by only 2% per year in the 2000s). The income growth and the rise in the shares of top groups

Figure 2.24: Evolution of Profits vs. Wages in Tax Declarations: 2007–2016



Notes: Cumulative growth of profits/dividends and wages/salaries identified in tax data. The former are directly calculated from the aggregate amounts of distributed dividends and profit withdrawals from both corporations and micro-enterprises. Wages/salaries of employees and self-employed workers are proxied from gross "taxable income", within which they are the main category.

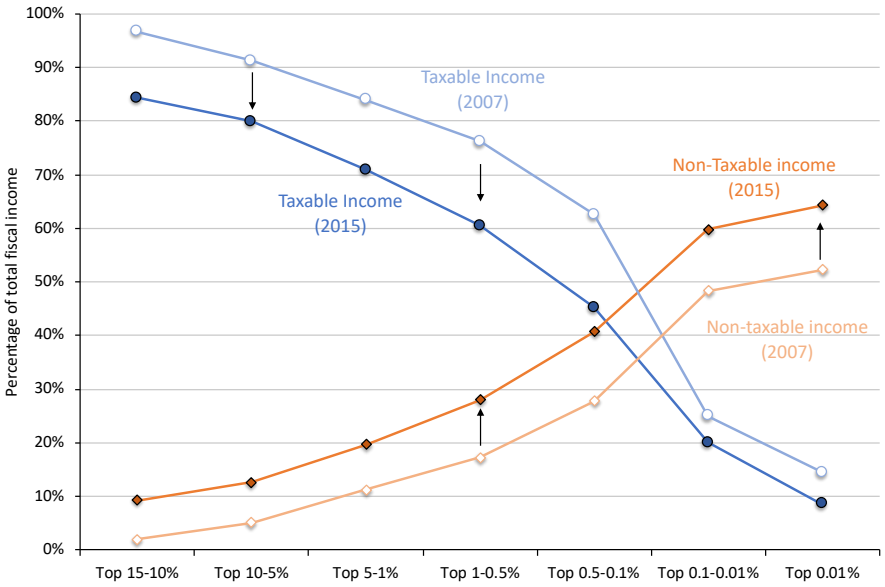
in national income (shown in sections 3.1 and 3.3) captures this. The extent to which private corporations chose to accumulate internally may well have been influenced by the uncertain economic and political landscape their country was traversing. It can also be related to the country's interest rate regime.

According to the national accounts, the growth of the retained earnings of non-financial corporations increasingly substituted interest-bearing borrowing, which is known to be a high burden on the financing costs of firms. Since 2003, Brazil's real policy interest rate (the *Selic* rate) was the fifth highest in the world, at an average of 6.3%, with the nominal lending rate averaging 13.25% (Weisbrot et al., 2016).[36] Like the government's minimum wage policy, the Central Bank's overnight lending rate acts as an anchor for other prices in financial markets. It is likewise, a policy choice variable. The exorbitant interest rates can partly explain the growth of retained earnings as a necessary financing channel for firms, but they also underpin the growing financialisation of the economy. From 2000 to 2008 the ratio of the stock of financial assets to the stock of fixed productive capital rose from 55% to 75% (Bruno et al., 2011).

According to the IBGE's national accounts, between 2010 and 2015, gross fixed capital formation of non-financial firms fell by 13% in real terms, while their net acquisition of

[36]Over 2003–2015, the four countries with a higher real policy rate were Gambia, Tajikistan, Belize and the Democratic Republic of Congo, according to IMF data. See Weisbrot et al. (2016).

Figure 2.25: Composition of Top Incomes in Brazil: 2007 vs. 2015



Notes: The graph shows the composition of incomes at the top of the distribution between taxable and non-taxable income in 2007 and 2015. As one moves up the distribution, the share of non-taxable income in total income increases. Apart from taxable income, the remaining share of total income is made up by exclusively taxed income, which also increase with income (reflecting the growing importance of interests and capital gains in top incomes).

financial assets grew by 40%. Overall, between 2004 and 2015 the real financial assets of the corporate sector grew by 6% per year, which was principally driven by the growth in bonds and loans (at 8% per year), and equities (at 5% per year). Therefore, real productive investment seems to have suffered the effects of high interest rates, from both the supply side (as a price for credit) and the demand side (as the basic return on financial assets). As the economy slowed down after 2014, firms piled more cash into lucrative financial markets, where they could avail of high-yielding, high-liquidity, risk-free government bonds and other securities – which offer protection from inflation and changes in exchange rates.[37] During periods of economic uncertainty and decline, the financialisation of corporate retained earnings seems to be soaking up resources that could otherwise be re-directed to the benefit of the Bottom 90% in the form of general investment spending in new employment opportunities, infrastructure and public services.

The combination of lower economic growth and higher interest rates since 2013 has also left the government in a bind. In 2015, the interest burden on the public debt absorbed 7.6% of GDP. Since 2012, the increase in this interest burden has accounted for about half of the rise in the federal government's deficit, which rose from 2.5% of GDP to 10.4%

[37]The growth of financial incomes in the personal distribution can be seen from the growth of "exclusively taxed incomes" in Figure 2.23. The growth of interests from financial applications (as registered in tax returns) grew at 8% per year from 2007 to 2016, the same as total exclusively taxed incomes.

in 2016 (Weisbrot et al., 2016). The politically charged nature surrounding the debates about public deficits, interest rates, inflation, and the government's overall "fiscal space" is underpinned by the substantial market and political power of the country's largest financial institutions.[38] The resilience of Brazil's high interest rates in the midst of country's worst economic crisis for three decades, highlights the policy preference of the monetary committee of the central bank, and the financial sector's indirect control of government policy-making. While a large federal expenditure package could have stimulated a recovery, exorbitant interest rates act as constraints on the imagined"fiscal space" of the government, and the very notions of such an expansionary policy. After the impeachment of President Rousseff in August 2016, the new interim administration, backed by Congress and the Senate, moved in the direction of "sound finance", passing a constitutional amendment (*PEC 241*) that freezes the real (inflation-adjusted) value of central government spending for 20 years (Couri and Bijos, 2016). This move will undoubtedly constrain the government's capacity to enact progressive expenditure policies, increasing the burden on collective wage bargaining and tax reforms to do the job. Given the pro-cyclical nature of the former (minimum wages evolve conditional on the GDP growth of the previous two years and on annual inflation rates) and the unaligned political coalitions necessary for the latter, Brazil's future inequality trajectory looks ominous.

5 Conclusion

While most studies on income inequality in Brazil have used either survey-based measures or (more recently) tax-based measures of inequality, this study combines these two sources, reconciling estimates with national accounts, to measure inequality over the past twenty years. In distinguishing between different income concepts and components, I contribute to the debate by clarifying the dynamics of inequality over the period. My results provide a sharp upward revision to the "official" survey-based estimates of inequality in Brazil. This confirms that surveys grossly underestimate the level of incomes at the top, and vastly over-estimates the income share of poorer groups. The notable result is the exceptionally large concentration of income at the top of the distribution, which seems to determine the shares of the rest of the population, including a Bottom 50% – who are highly dependent on government intervention in the economy, either in the form of minimum wages or cash transfers in the secondary distribution of income – and a "squeezed" Middle 40%. While the role of welfare cash transfers should not be diminished at the household level, their importance in the aggregate distribution is very limited due to their small share in the economy. Thus, greater onus has been on wage policy and between-group variations in income.

Income growth in Brazil has also been unequal, with the Bottom 50% making gains,

[38]The four largest banks in the country, two of which are public, account for 70% of the financial assets of the commercial banking sector (Weisbrot et al., 2016).

but not at the expense of top groups, who managed to further grow their share of total income since the domestic recession. Overall, economic elites still manage to capture disproportionate fractions of total growth due to their disproportionate share in total income. This shows that over the short-to-medium term, it is the share of income that matters more than its growth.

My results also show that labour income inequality declined, consistent with estimates from raw surveys (albeit at different levels). However, the change in labour income inequality and the distributional composition of income were such that it did little to compensate the persistent concentration of *total* income. Inequality of total income among the Bottom 90% decreased, influenced by the compression of labour incomes and earnings, and the lack of capital assets held outside the Top 10%. However, I also provide suggestive evidence that this decline could have been contaminated by income switching on the part of business owners. Thus, while certain policies can be linked to the decline in labour income and earnings inequality, their effects have been too small (or too contaminated by behavioural changes) to make large in-roads into distributional outcomes over the time-span analysed in this chapter. When measured in the context of the national income distribution, the effect of progressive social policies has been largely overridden by the inequality of capital, and the neglect of reform in certain areas, like taxation and monetary policy.

The current political turmoil in Brazil is the culmination of multiple factors: from attitudes towards corrupted political representatives, to the political aspects of the severe domestic recession. Yet, it also bears the mark of an economically divided country, where class divisions have arisen out of government-directed policies towards the poor, largely to the detriment to the upper-middle classes. Understanding the historical legacy of this distributional struggle and the relative resilience of economic elites to political and economic changes is a subject of notable importance. Chapter 3 pursues an investigation in this direction.

Chapter 3

Distributing Growth in a Context of Late-Development: New Evidence on Long-Run Inequality in Brazil (1926-2016)

Abstract: This chapter presents new evidence on the distribution of income and growth over the long-run in Brazil, combining survey, tax and national accounts data, alongside recently developed accounting frameworks and statistical techniques. Our results reveal the huge magnitude of income concentration and dispersion among the Brazilian population across ten decades, a persistence not seen anywhere else in the world. The Top 1% share has rarely deviated outside of the 20-30% bound and it has been entirely responsible for maintaining the Top 10% share stable around 55% since the 1970s. Top incomes have outgrown average incomes since the 1920s, and elites have successfully relocated the distributional conflict lower down the distribution – as one effectively between the poor and the "middle class". We also discuss the increasing irrelevance of surveys to adequately measure income growth and disparities in Brazil. Notable exceptions to the *status quo* can be found in the falling concentration of the 1942-1964 period, the poverty reduction of the 2000s, and in the striking decline in gender inequality since the 1970s. Our historical analysis explains the un-sustained egalitarian levelling over the century through an endogenous theory of institutional change, in which distributional imbalances related to "fundamentals" (land, capital, income, education) bring development into conflict with inflation and existing institutions. The inevitable crises that steer the direction of institutional change are closely correlated with income concentration, as we show in the Brazilian case. We identify the military *coup* of 1964 as the key point of inflection in Brazil's inequality history, whose legacy can still be perceived today.

1 Introduction

The optimistic belief that indefinite economic growth would lift all boats and placate social tensions has faded across most of the world.[1] In turn, distributional concerns have made a comeback both in academic and political circles. This was largely due to new research on inequality that revived the use of income tax data to shed new light on income concentration across countries over the last century (Piketty, 2003; Piketty and Saez, 2003; Atkinson and Piketty, 2007, 2010).

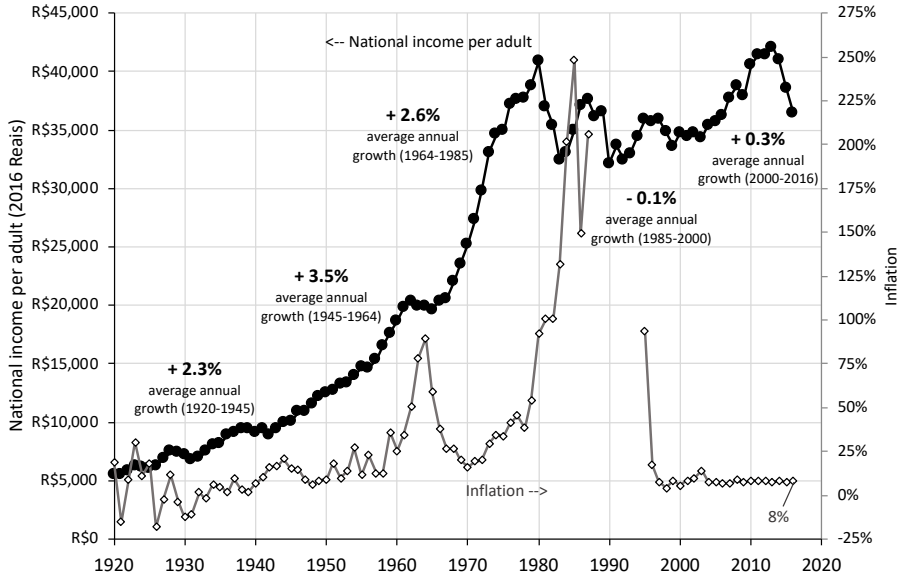
Yet, inequality analysis would not be fully consistent with macroeconomic development, and economic growth more specifically, until a group of these researchers proposed the construction of Distributional National Accounts (DINA) – a framework much more apt to study the distributional aspects of growth (Alvaredo et al., 2017). Along with the previous innovations, this work served as the foundation for the World Inequality Database, with developed-country case studies leading the way (Piketty et al., 2018; Garbinti et al., 2018).[2]

However, income inequality in most developed countries generally pales in comparison to underdeveloped countries – with the exception perhaps of the United States. Brazil is no different, providing a context of extreme inequality which is more often than not taken as an undesirable mark of underdevelopment. Unsurprisingly, the issue has always been hotly debated in the country, and seldom ignored. From Getúlio Vargas in the 1930s to Dilma Rousseff in the 2010s, innumerable government officials promised to address the so-called "social question", or, in contemporary terms, pursue inclusive growth. This was difficult to assess in the past due to data and conceptual limitations. With the birth of DINA and the availability of distinct historical and contemporary data sources for Brazil, as well as new statistical techniques, the "social question" over the long run of history can examined more accurately than ever before.

Brazil offers a unique opportunity to study the long-run evolution of inequality and its determinants. In particular, the arc of Brazilian history over the past ten decades shows periods of fast growth and stagnation; intense structural change and de-industrialisation; democracy and dictatorship; and so on. Figures 3.1 and 3.2 provide a glimpse of such changes, by showing, respectively, the evolution of average national income per adult alongside the inflation rate, and the sectoral composition of the economy. From 1920 to 1980, real average income grew 7.4 times (3.4% per year), while the manufacturing share of economy activity almost trebled. At the same time, the share of the labour force employed in agriculture fell from 70% to 30%, while that of industry rose from about 13% to 24%. Moreover, Brazil's average income represented 28% of that of the USA in 1950 (in USD PPP). This progressively rose to almost 50% in 1980. Over the period, prices grew by 18.1% per year, with inflation reaching 90% in 1980. After 1980, however, short growth spurts and severe recessions and inflationary crises led to an 11% fall in average income

[1] This chapter was written with Pedro Souza.
[2] https://wid.world/.

Figure 3.1: National Income Per Adult and Inflation in Brazil: 1920–2016



Notes: authors' calculations using data from IBGE. In 2016, average per adult national income corresponds to 18,257 USD PPP. Inflation refers to the annual price change as given by the GDP deflator. Estimates for 1988–1994 are omitted due to their extreme levels (the average rate was over 1,400%).

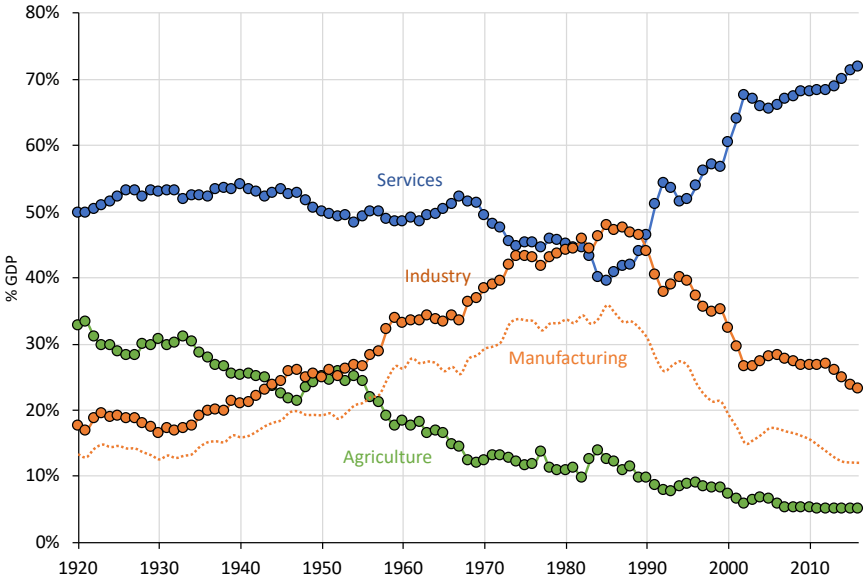
per adult (0.3% per year), and premature de-industrialisation – by 2010 industry's share of the labour force had fallen to 20%, while the share of services in GDP continued its upward rise, reaching 65%, from 46% in 1980 and 16% in 1920. By 2016 Brazil's per adult income had receded back to 27% of the US level, as inflation was restored to single digits in the late 1990s.[3] In short, Brazil has become much richer over the past ninety years, but this happened mostly during the heyday of import substitution industrialisation (ISI), which came crashing down in the 1990s.

In this paper we construct a new, long-run series of Distributional National Accounts (DINA) for Brazil from 1926 to 2016 to analyse the distributional consequences of such large-scale shifts. We combine harmonised household survey data, income tax tabulations and national accounts data to compute income shares and inequality measures for the full adult population and for top fractiles across nine decades. We present both pre-tax national and fiscal income series, as well as a systematic discussion of the shortcomings of survey measures to track the development of growth in the country.

Hence, our paper provides a novel contribution to the well-established body of research of Latin American inequality. In Brazil, in particular, empirical research on income inequality blossomed in the early 1970s and for decades relied almost exclusively on household surveys for data (e.g., Hoffmann and Duarte (1972); Fishlow (1972); Langoni

[3]Sectoral labour force data is from (IBGE, 2006) and ipeadata (`http://www.ipeadata.gov.br/Default.aspx`). Data for the Brazil–USA income gap is from `https://wid.world/`.

Figure 3.2: Gross Value Added by Economic Activity in Brazil: 1920-2016



Notes: estimates are sourced from IBGE (2006, 2017).

(1973)). More recently, researchers turned to income tax data to overcome the known underreporting of top incomes in surveys. Medeiros et al. (2015a,b) inaugurated this trend by showing that surveys greatly exagerated the alleged fall in inequality in the 2000s, and Souza and Medeiros (2015), Souza (2016, 2018) and Morgan (2015) estimated top income shares in the long-run. Afterwards, many others reanalysed the same data, often focusing on the distributional effects of the federal income tax, like Fernandes (2016), Castro and Bugarin (2017) and Gobetti and Orair (2017a). More significantly, Morgan (2017) was the first to calculate full-fledged DINA series for 2001-2015, going beyond "fiscal income" – an important step in itself, insofar as DINA allows us to distribute all income flows as measured in the national accounts (see Section 3.1), and better equips us to address pressing questions about economic power in society (see chapter 2).

We also offer a historical analysis of the growth-inequality nexus in Brazil, which sheds new light on the endogeneity of institutional change in a context of late-development. We pit institutional theories against structural-market theories of inequality, and highlight how economic growth and distributional shifts are intimately tied to political and institutional disruptions. In a broad sweep, the Great Depression was the proverbial straw that broke the Old Republic's back, and the early 1960s stagnation amid political polarisation gave way to a military dictatorship, which in turn had its own collapse accelerated by the early 1980s contraction. Recently, the severe recession of 2014-2016 brought the long-standing PT administration to a sudden halt.

We argue that deliberate and rapid industrialisation in late-developing countries like

Brazil inevitably arrives at distributional bottlenecks, causing institutional shifts that are capable of fundamentally altering the path of inequality. These bottlenecks are underpinned by social conflict over the institutional framework needed to foster further development and promote higher living standards for the large mass of the population. Therefore, the political outcomes of such recurrent crises have huge consequences for inequality. We show that this been the case in Brazil – a rapidly developing country with a highly segmented social structure. Without the unifying effects of wars or mass revolutions to fall back on, Brazilian distributional history has been a secular tussle between powerful social actors and political representatives over how to distribute the proceeds of development *for continued development*.

The remainder of the paper is structured as follows: section 2 surveys the the key literature on economic development and income inequality; section 3 discusses concepts, sources and methods we use to estimate DINA; section 4 presents our main results for income shares, the distribution of growth and gender inequality, including international comparisons; and Section 5 examines the historical context in more detail and lays out our interpretation of the results. Appendix B provides further methodological information.

2 Inequality and Development

2.1 Theoretical Perspectives

In theorising about the co-variation of inequality and economic development, multiple generations of social scientists have recognised that distributional issues are intimately tied with the development of a society's productive forces. The impact of adopted modes of production on the relative prices of factors of production (and the classes comprising them) has been generally understood to operate through at least one of two principle variables: cyclical/secular economic forces, and endogenously-determined institutions.

We can identify two broad types of theories of inequality and development, according to their distributive determinants. On the one hand, there are theories that bring to the foreground economic/market forces. Classical and Neoclassical theories of distribution are of this sort for the most part. They postulate that incomes are mainly supply-determined through changes in savings, demographics and the supply of natural resources. The former theories, in the Ricardian tradition, regard distributional shifts as ultimately tied to the diminishing returns of land, where wages are fixed and profits are the residual left over from the difference between the marginal product of labour (produce minus rent) and wages. As capital accumulates, so does the demand for labour, pulling population growth with it over lesser quantities of fertile land, such that rent absorbs an ever greater share of the marginal product of labour (Kaldor, 1955). In a similar vein, Classical-inspired dualist (traditional-modern sector) theories of the twentieth century mostly abstracted from political and social forces to trace productivity growth to savings and capital accumulation. This accumulation

is not subject to population limits – with an unlimited supply of labour available in the traditional sector – and alters income distribution in favor of the high-saving class (Lewis, 1954; Kuznets, 1955). The extent to which abundant, cheap labour in the traditional sector holds down wages in the modern sector is a point of contention in these theories. Lewis (1954) argues that economic forces (population growth, immigration, capital exports) prevent increases in wages, while Kuznets (1955) assumed that the higher productivity in the modern sector pulls its wage-earners' remuneration above levels observed in the traditional sector. The notable assumptions in these two brands of theories are about the fixation of wages at or near subsistence levels, the between-sector inequality differences and whether modern-sector wages rise with capitalist productivity.

Like most of their Classical counterparts, the Neoclassical theories (set in motion by Walras and Wicksell in the late 19$^{\text{th}}$ century, and embedded in most mainstream theories of growth) depend on economic/market forces. However, they are much less precise in identifying the locus of distributional differences and change. For these theories, income distribution is determined essentially by the market's efficient allocation of resources, which equilibrates economic forces operating in different directions. These forces ensure that prices (profit margins and wages) are determined by the optimal amount of capital used relative to the available labour ("marginal productivity" variants) or through competitive forces being such as to ensure that profits are not too high or too low ("degree of monopoly" variants). Thus, productivity (enhanced by physical and human capital accumulation), labour-force growth and competition are made to determine income distribution and investment in the short-term and growth in the long-run. However, problems of circularity arise when trying to insert some form of historical time into the picture. When doing so, it becomes clear that these theories (particularly the marginal productivity variants) amount to stating that present factor prices are determined by past factor prices, which are themselves determined by the marginal rate of substitution between capital and labour. Yet, the marginal rate of substitution "can only be determined once the rate of profit and the rate of wages are already known" (Kaldor (1955), p. 91). Thus, it would seem that these theories do not say anything meaningful about distribution in a growing economy.

On the other hand, there are theories of inequality and development that lend more weight to the institutional determinants of distribution and its change over time. Following the Classical tradition, the Marxian theory recognizes the importance of economic forces for the distribution of income between classes (changes in prices and quantities linked to capital accumulation, technology and competition). But Marxian analysis equally focuses on institutions; mainly as power relations expressed through ownership of the means of production that defines profit-recipients as capitalists/employers and wage-recipients as workers/employees. The mode of production defines relations of production, which are embedded in property rights and labour contracts. Supporting these relations are society's legal and political "superstructure" (the state) and collective social actors (classes), which jointly administer conflicts over the direction of technological change and the distribution

of its proceeds. For Marx, distributional shifts come from in-built characteristics of the dominant mode of production. In Capitalism a "reserve army" of workers keeps wages low, but capital accumulation is hypothesised to bring forth an impending crisis in the profit rate due to labour scarcities during full employment (competitive-stage) and lower demand from increased displacement of workers by fixed capital first, and then by lower market compulsions to re-invest profits as industry becomes concentrated into large-scale producers (monopoly-stage). Importantly, the crises that such forces promote have to be mediated institutionally (Marx, 1990; Kaldor, 1955).[4]

While often overlooked, Kuznets' downward inequality cycle is fully determined by institutional factors. While economic forces mostly explain the changes in the upward cycle of his secular 'inequality curve', legal and political interventions (taxation, inflation, price controls, etc.) limit the degree to which the accumulation of savings and property can be inequality-enhancing in the long run. According to the author, they reflect skepticism on the part of society on the long-term usefulness of large income disparities: "This view is a vital force that would operate in democratic societies even if there were no other counteracting factors. This should be borne in mind in connection with changes in this view even in developed countries, which result from the process of growth and constitute a re-evaluation of the need for income inequalities as a source of savings for economic growth. The result of such changes would be an increasing pressure of legal and political decisions on upper-income shares – increasing as a country moves to higher economic levels" (Kuznets (1955), p. 9).

Although not conventionally applied as a distributional theory, the Keynesian theory can be equally used as a theory of distribution, as it can as a theory of output and employment – the latter being what it was initially conceived for (Kaldor, 1955).[5] In essence, the application of the Keynesian theory to the problem of distribution involves the combination of economic forces and institutions; that is, of quantities, prices and conventions. The schema can be sketched out as follows. The quantities of factors of production are determined by the circular flow of multiplicative relations in a monetary economy. Income (profits and wages) are dependent on expenditures (consumption and

[4]For the negative relation between fixed capital accumulation and the profit rate to hold during the competitive or monopoly stages of capitalism, both wages and the size of production units must be subject to upward pressure. This would depend on the bargaining power and collective organisation of workers, on the one hand, and on the prevalence of some form of anti-trust legislation on the other. The former point on wages was emphasised by Joan Robinson in *An Essay in Marxian Economics*, 1942 (Kaldor, 1955).

[5]According to Nicholas Kaldor, "there is evidence that in its early stages, Keynes' own thinking tended to develop in this direction [towards the problem of distribution]", particularly in the *Treatise on Money* (Kaldor (1955), p. 83). It could also be argued that elements of a theory of development applied to inequality was present in the *Economic Consequences of the Peace* (see Keynes (2011), p. 12-13, and further analysis below). According to Kaldor, the reason why Keynes never developed his multiplier-analysis into a theory of distribution was the idea "that inflationary and deflationary tendencies could best be analysed in terms of resulting changes in output and employment, rather than in their effects on relative prices" (Kaldor, 1955).

investment) among *given* classes of capitalists and workers, who save and contribute to expansion in unequal proportions. Investment (negative savings) is the key independent variable in the process, given the constancy of class structures and saving propensities. The distribution of income (the prices of factors) is determined by the relationship between wages and the business mark-up (price including profit, after wages are paid). Keynesian analysis thus theorizes about the direction that certain variables have on others as the economy evolves. However, the share of the total product captured by different individuals (the share going to profits and wages, and the within- and overall-shares of income) is one of *convention*. At its core, distribution is a moral dilemma that evolves with ideas, ethical sentiments and political change. The conventions created from these three factors not only affect the distribution of income between broad classes, but also get passed through to other dimensions of social inequality, such as gender and racial inequality.

2.2 The Primacy of Conventions

Analysed sequentially and historically, income distribution seems to be "given" by politically- and conventionally-determined ownership and control of key resources for development, namely land and capital, whose income flows are the proceeds of a country's economic expansion. Classical economists, and economists in the Keynesian tradition, posited that saving propensities of different income earners/classes/sectors are not equal. Thus variations in income distribution bring about changes in total savings and aggregate demand via consumption/investment decisions, which may impinge upon certain market forces, like competition, or certain institutions such as government price controls.

Fundamentally then, conventions about *how* a society can accumulate and progress materially get transmitted into institutions that coordinate expectations about the future to ensure economic and social stability (Blyth, 2002). For Keynes, given conditions of radical uncertainty about the future (a key component of his entire edifice), economic agents must rely on certain rules of thumb, which are, in essence, about "*conventional* judgement" (Keynes (1964), p. 214).[6] These conventions, in the Keynesian sense, "are the shared ideas about how the economy *should* work" (Blyth (2002), p. 43). It is these shared ideas that underpin the balance between institutional stability and instability over the long run. Ideas help to construct institutions, which once in place reinforce the ideas that build them through conventions (or "national narratives") (ibid). For instance, the extent to which capital owners may claim as theirs the surplus of production and decide what to do with it has important conventional connotations, tied to particular institutions in place at particular points in time. And how the distribution of income may be tilted in their favour for being perceived to have the responsibility to do so, may well depend on

[6]This is nowhere more evident than in chapter 12 of his *General Theory*. His propositions about "animal spirits" and "beauty contests" depend upon conformed beliefs "with the majority or average."

how development and distribution are understood to co-vary by economic agents.

The most powerful theoretical contribution in this regard, being most common across all theories, was the idea that directing income to individuals with the highest saving propensities (effectively making the distribution more unequal given that high earners are also high savers) was the most effective means towards economic improvement, as accumulation was thought to be intimately tied to saving. However, the crucial aspect of this theory is that a surplus can either become a leakage or a re-insertion into the circular flow of income, and that future savings are determined by current re-insertions in the form of investment. The historical legitimacy of such a system depends precisely on the onus on capital owners to contribute enough of their politically-determined surplus to develop society's productive forces (Keynes, 2011; Palma, 2011).

2.3 Problems of Late-Development

Within the more dynamic cohort of theories, there are those focused on economic expansion in general (Classical and Keynesian theories) and those with an applied focused on the problems of late development (Keynesian-Structuralist tradition). The latter focused on the insertion of less-developed countries into the world economic system, which is divided into an advanced "centre" and a lagging "periphery", subject to different productive and sectoral structures. This system has been politically shaped both through direct means (colonial exploitation) and indirect means (commercial exploitation), and maintained through conventions about the course of development held by economic agents in both regions. The relationships between these two world economic regions (and their policymakers) are crucial in understanding both between-country and within-country evolutions in growth and inequality. Indeed, many of the periphery's problems in converging with the advanced capitalist economies are at least partly due to its commercial and institutional interaction with the centre, through trade agreements and pricing policies. This interaction has resulted in peripheral economies with fragmented and highly specialised productive structures vis-á-vis the more homogeneously diverse ones of the centre (Blankenburg and Palma, 2012).[7] It has also resulted in greater dependence on foreign multinational corporations for domestic industrialisation. The specialisation in (mainly) primary products results in a dynamic of exporting low-income-elasticity-of-demand goods to the centre in return for high income elasticity goods from the centre. This naturally puts a "material limit to the rate of growth of real income in the periphery relative to that of the centre" (Blankenburg and Palma (2012), p. 140), unless targeted government intervention changes this trend. These dynamics are linked to the distribution in various ways, from the degree to which the ownership of resources (e.g. land) in the primary-agricultural sector is concentrated, to the

[7]The early pioneers of this tradition of thinking in Latin America were Raul Prebisch, Aníbal Pinto and Celso Furtado; the latter being responsible for the spread of its ideas in Brazil. All three were associated to the United Nations Economic Commission for Latin America and the Caribbean (ECLAC).

extent to which an import substitution strategy for industrialisation redistributes income to expand the domestic market.

Due to their qualitatively distinct problem-set, the tolerance for income disparities as a necessary driver of development in late-developing countries may be lower than in early-modernising nations. Indeed in the advanced nations, a mix of relative social and cultural homogeneity (ethnic, religious), shared historical experiences (wars and revolutions) and common institutional features allowed for the so-called "tunnel effect" (tolerance of inequality based on expected future mobility given current mobility of their peers) to be very operational.[8] By contrast, late-developing countries are typically more segmented in their social and regional structure and vary in the intensity of their shared historical experiences. Thus, "highly segmented societies will or should eschew strategies of development that are politically feasible elsewhere because of the availability of the tunnel effect" (Hirschman and Rothschild, 1973, p. 554). In these societies standard capitalist development may well "require" a higher degree of coercion and more centralised economic planning, which are themselves prone to diminish the tolerance for income disparities. For different reasons, centralised-decision-making economic systems are forced to be either egalitarian or in-egalitarian, depending on their *raison dêtre*. Thus, given existing social, psychological or political structures, the operation of the tunnel effect can allow for growth and equity in the income distribution to be pursued either sequentially (growth prior to equalisation) or simultaneously (Hirschman and Rothschild, 1973).[9]

2.4 Brazilian Perspectives

In light of the above expositions, Brazil is an interesting case study for numerous reasons. First, we are in the context of a large land- and labour-abundant peripheral country, engaged in late capitalist development, at least from the 1930s. This brings forth qualitatively and quantitatively distinct challenges, compared to those experienced by the advanced "centre" of the world economy. The pass-through of production patterns to the distribution of income during the development cycle (via saving/investment propensities, income elasticity of demand, cost-push inflation, sectoral productivity and employment, educational compositions, government policy), mean that the nature, sequencing and timing of sector-specific activities are crucial in determining the evolution of inequality during development. And these can operate differently in the less-developed periphery, given its dynamic interaction with the advanced capitalist centre. The significant variation in Brazil's

[8] Emphasis is often place on the unifying capacity of particular historical experiences shared by a community, which help to construct narratives about national identities. Two emblematic cases are the "born equal" heritage of the USA and Sweden's *Sonderweg* egalitarianism (Hirschman and Rothschild, 1973; Bengtsson, 2017) While, the former facilitated the acceptance of often large economic disparities, the latter shaped a particular intolerance to high inequality.

[9] However, the appropriate solution for each country is hard to know ex-ante, so that "only development itself will tell" (Hirschman and Rothschild (1973), p. 561). This fact places the responsibility on retrospective historical analysis.

economic growth and sectoral structure over time, as evidenced from Figures 3.1 and 3.2, makes it a particularly salient case to analyse.

Furthermore, Brazil forms part of a region historically characterised by high levels of inequality relative to other areas of the world. Economic inequality has followed a long history of social distinction, similar to other parts of the world, but in Brazil it has been particularly acute. In the colonial period (c.1500-1822) and the subsequent independent monarchical period (1822-1889), the law clearly instituted social and racial inequality. After gaining independence from Portugal in 1822, only domestically born, literate, male Catholics with the requisite income could vote in Brazil. An electoral law of 1881 extended the suffrage to non-Catholics and naturalised citizens. Yet, it still excluded women, illiterates and slaves – the latter two are estimated to represent about 80% and 30-35% of the population respectively – as well as the literate poor (Love, 1970; Finlay, 1980). The institution of slavery was abolished in 1888 (the last country in the Western hemisphere to do so), paving the way for the Federal republican constitution of 1891. The latter removed the income threshold on voting, but nevertheless maintained the gender requirement (until 1934) and the literacy requirement, which would persist until 1988. Thus, much of the poor continued to be excluded from the political process for another century.[10]

But as evidenced by Williamson (2015), it is "Latin America's 20^{th} century inequality history which is unique, not its colonial history, nor its early republican experience, nor its *belle époque*" (ibid). According to the author, the principal reason why Latin American countries are so unequal relative to other regions is that they largely missed out on any sustained mid-century egalitarian movement. But "why did Latin America miss the Great 20^{th} Century Egalitarian Leveling?" (ibid). This question took centre stage in the Brazilian context, with a famous, yet controversial, distributional debate occurring in the 1970s among Brazilian economists (involving academic economists with different degrees of association to the military government), but also attracting foreign contributors (see Taylor et al. (1980)). The subject matter was about the extent to which inequality had increased between the early 1960s and the early 1970s (a period of marked economic and political change), as well as the reasons for the observed changes. Just preceding this debate was another important debate between Brazilian economists over the degree of stagnation in the economy and its links to income inequality (Furtado, 1969; Serra and Tavares, 2000). By the 1980s Brazil had become a focal point in the emerging world for models of growth and distribution. Interest on inequality and development would re-surface again in the 2000s, after the stabilisation of the macroeconomy and the resurgence of growth. At the heart of these debates is the institutional production and use of statistics by each faction.

We henceforth turn to examine how unequal the distribution of income really was in Latin America's leading economy, and whether it missed out on the egalitarian levelling

[10]The illiteracy rate would not fall below 50% until the 1950s. By the year of the condition was repealed, in 1988, the rate was still 20% (ipeadata; IBGE (1952)).

since the 1920s, considering the interpretations previously advanced.

3 The Brazilian Case: Concepts, Data and Methods

A long list of studies have used household surveys over the years to analyse different dimensions of economic inequality in Brazil. A much more reduced set of studies have used income tax data for the same intentions. This contrast is largely the result of the restricted availability and complex reconciliation of both data sources. This paper is not the first to use tax records to study inequality (see chapter 2 and Mortara (1949a,b); Langoni (1973); Medeiros et al. (2015a,b); Morgan (2015); Souza and Medeiros (2015); Souza (2016, 2018)), nor is it the very first to seek a combination with survey data (see chapter 2 and Souza (2016, 2018); Medeiros et al. (2015b, 2018)).[11] The methodological novelties of our contribution compared to much of the studies just cited, is that we provide a more precise and comprehensive view on total income inequality across multiple data sources (using the same basic data sources and national accounting framework as chapter 2) over a longer time horizon. The following sections explain the income concepts we use, our data and estimation method.[12]

3.1 Income Concepts

Our ultimate goal is to distribute net national income from the national accounts. This is the total income that flows annually to national residents of Brazil (comprised of various institutional sectors – households, non-profit institutions, corporations and the government), after accounting for depreciation of fixed capital.

Research on income inequality usually relies on more limited income concepts, such as the "disposable" (i.e. after direct taxes and transfers) or the "gross" (pre-tax) income of households, as it is usually measured in nationally representative surveys or in some types of administrative data. Our focus on net national income accounts for income flows that may escape concepts tied to particular data sources, if analysed in isolation. Moreover, the degree and magnitude of these "missing" flows may depend on both country-specific and time-specific reasons, such as survey sampling error (due small sample size), survey

[11] Mortara (1949b,a) was the first scholar to make use of personal income tax returns in Brazil, even applying the Pareto interpolation to the tax tabulations. However, his contribution did not spur further studies until the 1970s, when scholars with ties to the military dictatorship, such as Kingston and Kingston (1972) and Langoni (1973), also relied on income tax data to try to push for more benign views of the rise in inequality in the 1960s.

[12] The content of these sections are very similar to the ones presented in the methodology part of chapter 2. Their difference lies in the presentation of data sources and data treatment for the years prior to 1995. Readers who are thus familiar with chapter 2 can focus their attention on descriptions for the 1926-1994 data and the estimation of inequality over this period.

non-sampling error (due to differences in response rates), tax laws and incentives, etc. Thus, to minimise the aggregate bias related to income coverage in our temporal distributional series, and to facilitate international comparisons, we target a concept of income that is broadly homogeneous across regions and over time.[13]

Adding the missing flows into the picture means that we must distribute all the income, attributed to different institutional sectors, to the unit of observation that interests us. In our case, as is habitual in much of the literature, our preferred unit of analysis is individuals residing in households. We choose to focus on individual adults (aged 20 and over) for reasons to do with the nature of income reported in tax data, which we regard as the benchmark data source for higher incomes.[14]

While we focus primarily on the distribution of national income, we do not neglect the distribution of other concepts of income that are products of different data sources. In our case we exploit surveys and personal income tax data, alongside national accounts. In doing so we wish to offer a multifaceted view on the distribution of income in Brazil, a country that, as highlighted, has drawn much attention and research to the topic. Above all, we aim to provide an improved assessment of "gross" income inequality to what has be done in the past.

For the Brazilian case, as is common to researchers of most less-developed countries and many developed countries, we do not have access to administrative microdata, to which information from surveys can be appended. Rather, we avail of tabulations of fiscal income that present (at a minimum level of information) the number of declarations and assessed income per income bracket. In less-developed countries in particular, the population coverage in the tabulations is often substantially less than the total adult population. In such a case we are obliged to proceed the other way round: rather than incorporating survey information into the tax data, we need to incorporate tax information into the survey data.

From the surveys, we can first estimate a distributional series of "survey income". Combining the survey data with tax data to correct for top incomes, we can then compute a series of "fiscal income". The income captured in these two series covers labour income, mixed income and capital income. More precisely this includes wages and salaries, pensions, self-employment income, interests, rents, distributed business profits and dividends, and capital gains. It thus corresponds to *pre-tax post-replacement fiscal income*, i.e. income received by individuals before personal income taxes, employee and self-employed social contributions (but after employer contributions), and legal deductions, but after accounting

[13]Evidently, national accounting concepts like GDP and national income are subject to their own measurement difficulties and temporal changes. But at least these latter changes are constant across countries, to the best of their efforts. Measurement issues can vary substantially, but they should be no worse than the issues that arise with national surveys or tax data. At the very least national accounts provide an agreed comparable framework for the dissection of the national economy.

[14]The choice of adults aged 20+ also facilitates international comparisons, being the age-cuff of for adults used by the United Nations; see Alvaredo et al. (2017).

for social security benefits in cash (unemployment insurance and social security pensions). All these items are included or excluded in order to make the income in the surveys consistent with the definition of income in the personal income tax declarations.

“Fiscal income" is distinguishable from “national income" insofar as it only concerns distributed income received by physical persons that is assessed by the tax office for fiscal purposes. It should also be distinguished from “taxable income", which is the income that is ultimately taxed after the application of legal deductions. Some components of income may have to be declared on the tax returns but are not ultimately “taxable”.[15] Pre-tax fiscal income constructed in this way can also be related to an equivalent total from national accounts.

The next stage involves moving from fiscal income to total “personal income". This implies that we factor in flows of income appearing in national accounts that (1) get attributed to households but are not included in fiscal income, such as imputed rents, investment income attributable to insurance and pension funds and social contributions made by employees and self-employed workers; and (2) are not attributed to households, but rather to corporations, such as undistributed corporate profits (i.e. the net primary income of corporations). Finally, moving from personal income to "national income" means that we must account for incomes that flow to the government rather than to households, namely government net capital income (government share of undistributed corporate profits, net interests and other incomes), social insurance surpluses and net production taxes. More precisely, total pre-tax national income can be computed as follows:

Total pre-tax fiscal income[16]
− Social contributions (D61, S14)
+ Imputed rent for owner-occupiers
+ Investment income attributable to insurance policyholders (D441, S14)
+ Investment income payable to pension entitlements (D442, S14)
+ Household/NPISH component of pre-tax undistributed corporate profits (B5n, S11+S12)
= Total pre-tax personal income
+ Government factor income[17]
+ Pension and other social insurance surplus[18]
+ Net production taxes received by the government (D2-D3, S13)
= **Total pre-tax national income (DINA)**

[15]This may vary with countries. In the case of Brazil it is explicit, as the tax declarations include a section for declaring non-taxable incomes.

[16]This can vary by country, but usually involves the sum of salaries + gross operating surplus + gross mixed income + gross interests + dividends + social security cash benefits + gross current transfers - imputed rent - capital depreciation of households.

[17]This equals net property income received by the government (D4n, S13).

[18]This equals pension and other social contributions (D61, S14) − pension and other social insurance benefits (D621+D622, S14).

It may seem questionable to impute income flows to individuals, which are not directly received during the year. A case in point is the undistributed profits of domestically-owned corporations. Under conventional practice, such retained earnings in "legal persons" (i.e. corporations) would rarely be considered income. Yet they are considered as such by the national accounts. Thus, these income flows, some of which constitute the build-up of assets, must ultimately have beneficial owners, which either exercise direct control over the flows or have future entitlement to them. But more importantly, excluding these "missing" flows introduces arbitrary biases in distributional series between countries, which are the result of between-country differences in legislation. In addition these biases can be time-varying. In the example of corporate profits, the extent to which they are distributed or not may well depend on incentives derived from country-specific tax laws or corporate governance legislation, which can change over time. And these decisions have important effects for the inequality we wish to measure.

Thus, in order to avoid (to the extent possible) producing "statistical artefacts", we prefer to concentrate on the distribution of national income. However, we shall compare this comprehensive series with the distribution of other income concepts, namely of raw survey income and total fiscal income, in order show how inequality can change between them. Furthermore, the discrepancies between the three series can have different implications for economic analysis and public policy.[19]

3.2 Data Sources

For the estimation of income inequality and growth incidence in Brazil we exploit three data sources – federal personal income tax declarations, household survey microdata, and national accounts data.

The Brazilian income tax was created in 1922. The Brazilian Tax Agency has never granted access to income tax microdata to third-party researchers, but fortunately there are numerous publicly available tabulations for federal personal income tax declarations (DIRPF) from 1926 onwards. We rely on these tabulations, which mostly come from official reports by the national statistics institute (IBGE) in the years up to 1960 and by the tax authority itself after 1960. For some individual years, when data from "official" sources were not available, we exploit tabulations used by independent authors in books or research papers (see Table B.1 in Appendix B). In total we have tax data for 73 years over the 1926-2016 period.

[19] While much economic analysis is interested in individual or household monetary "welfare" (either from gross or disposable income), questions of economic *security*, *capability* or even *efficiency*, more generally, should not be overlooked. What may apply to one individual or a group of individuals may not apply in the aggregate. Ultimately, the pass-through of incomes from the pre-tax national distribution to the pockets of individuals is mediated institutionally, by corporate legislation, fiscal policy and social conventions, etc. To better understand the entire process it is necessary to begin with the pre-tax national distribution. This conceptual analysis can be interpreted as bridging the gap between the micro focus on individual welfare and the macro focus on economic aggregates.

A complex feature of the tax statistics is, that for numerous years, they present tabulations for distinct income concepts to the one we are interested in exploiting from them, namely total pre-tax fiscal income (as highlighted in Table A.1). Not only does the concept of reported income change across years, but so does the concept of the ranking income across brackets and the geographical coverage of the tabulation. We explain below how we get around these issues.

The survey data corresponds to the microfiles of the *Pesquisa Nacional por Amostra de Domicílios* (PNAD), a large multi-purpose nationally representative survey run by the national statistics institute (IBGE). For our income analysis we use the individual-level microfiles available between 1976 and 2015, which we extract directly from the IBGE website.[20] The number of income variables rose over time from 8 in 1976 to 14 from 1992 onwards. As usual, the PNAD underestimates top incomes, especially business and investment incomes (see Hoffmann (1988); Souza (2015) and chapter 2), but it is otherwise known for being a high quality dataset. It has been the IBGE's flagship household survey for decades and the major source of income data for research on poverty and inequality.

The survey was conducted annually, except in Census years (1980, 1991, 2000 and 2010) and in 1994 (due to budget cuts).[21] The data are nationally representative from 2004 onwards. Previous years do not include rural areas of six northern states (Rondônia, Acre, Amazonas, Roraima, Pará, and Amapá). Moreover, the data for 1976-1979 also excludes rural areas in center-western states (Mato Grosso, Mato Grosso do Sul, Goiás and Tocantins, which was still part of Goiás). According to the 1980 and 1991 Censuses and the recent PNADs, this amounts to excluding 3-4% of the population in the 1970s and about 2% between 1981 and 1990.

Finally, we also make use of national account estimates. Brazil has a long history of official national accounting, dating back to the 1940s. Our objective is to link all income aggregates, whenever possible, to the national account aggregates. Thus, for our purpose of building DINA, we use the most up-to-date national account statistics, which generally come from the IBGE, for most years (IBGE (2000; 2006; 2017)), or from the Fundação Getúlio Vargas (FGV) for some selected earlier years, whenever statistics from the IBGE were not available (FGV, 1962). Comprehensive and integrated national accounts are available up to the year 2015. For 2016 we only possess totals of the main aggregates. Thus, we extrapolate totals for all sub-components maintaining the share of each component in national income for this latter year.[22]

[20] https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html?=&t=downloads.

[21] The PNAD was discontinued in 2015 as IBGE transitioned to PNAD Contínua (PNADC), a quarterly survey with a rotating panel similar to the United States's Current Population Survey (CPS). Both the sampling design and the questionnaire were overhauled, so the PNADC results are not directly comparable to the historical PNAD series for most variables.

[22] Section 3 of Appendix B explains the historical evolution of national accounting in Brazil.

3.3 Adjustments to Tax and Survey Data

The preparation of the data for building Distributional National Accounts (DINA) needs to account for the different degrees of income coverage in each distributional source. Thus, we have to harmonise tax and survey data to adhere to the same income concept before combining them with other information to compute our "national income" (that is, *pre-tax post-replacement national income*) series. While this income concept can, by definition, be constructed from national accounts, it cannot be directly composed from our two main sources of distributional data. This is because the tax tabulations have to be adjusted for missing information and other inconsistencies appearing over time. And surveys lack direct information on certain necessary components, such as imputed rents and social contributions paid out of wages/salaries.

We describe below the major procedures applied to both data sources. More detailed information can be found in the Appendix.

3.3.1 Tax Data

The raw tax tabulations require four kinds of imputations (see Section A.4.1 in the Appendix). First, the published statistics are not nationally representative until 1943 (and in 1966), so we scale-up the published totals to produce nationwide numbers by using information either on the regional distribution of taxpayers or of the income tax revenue. We validate this approach by comparing scaled-up regional estimates with national figures in years when both kinds of tabulations were available (1945-1947), finding that the results are very close. Second, we rely on neighbouring years with complete information to impute the number of tax returns per income brackets in 1963 and 1964. Third, most tabulations until the mid-1960s presented information only on "net taxable incomes", which have to be transformed into "gross taxable incomes" via the imputation of personal allowances and schedular deductions. Again, we rely on neighbouring years with complete information to do this. Fourth, we use published and/or imputed data on deductions to impute fixed capital consumption of unincorporated businesses. Additionally, we also have to disentangle the data for 1926 from the tabulations for 1927, which conflated incomes from both years.

Albeit extensive, these adjustments are robust and quite uncontroversial, and they either replicate or refine procedures widely employed in the existing literature (see chapter 2 and Medeiros et al. (2015a); Morgan (2015); Souza (2016); Morgan (2017); Souza (2018)).

3.3.2 Survey Data

The PNAD microdata has information on gross monthly incomes by source, such as labour earnings, pensions, and so on, with some minor changes over time. Overall, the data is quite reliable, though notably thin on property incomes as compared to the tax data (Hoffmann (1988); Souza (2015) and chapter 2). In order to reconcile the PNAD income

concept with our definition of "fiscal income", we proceed in three steps: imputations, reclassification and annualisation.

The imputations are needed because the PNAD does not collect information on infrequent incomes that have to be reported on tax returns. These incomes correspond mostly to benefits of formal employees defined by labour laws, such as unemployment insurance, FGTS, *abono salarial*, the holiday bonus and the 13^{th} salary.[23] Fortunately, these benefits can be easily calculated by applying labour laws to standard PNAD variables such as employment status (see section A.4.3 in the Appendix for details).

The reclassification step assigns observed and imputed incomes to three main "fiscal income" categories: labour, capital and mixed. labour incomes encompass the labour market earnings of employees, pensions and survivors' benefits, a fraction of employers' earnings, and all imputed and/or observed labour-related benefits. In our benchmark scenario, the labour fraction of employers' earnings is equal to income up to the exempted threshold defined by the personal income tax. Excess incomes above that threshold were classified as "capital", alongside financial (that is, "other incomes" greater than the current minimum wage) and rental incomes. In turn, mixed incomes are the earnings from the main occupation for self-employed individuals with no employees.

Finally, the annualisation factors take into account both monthly inflation rates and the reference months of the survey (typically, September).[24] Our factors range from 9.30 in 1992 to 12.01 in 1998. As expected, all multipliers after the macroeconomic stabilisation in 1995 are very close to 12.

Given we are interested in "national income", we also use surveys to estimate the distribution of non-fiscal incomes like imputed rents and Social Security contributions. The latter were straightforward: analogously to the imputation of labour benefits, we apply prevailing rates to survey variables such as earnings and employment status, with the aid of a few simplifying assumptions. For imputed rents, we adopt the standard two-step approach in the literature (Ferreira et al., 2003; Morais and Cruz, 2003; Soares, 2017): first, we estimated hedonic rental price regression models for rented dwellings; then we apply the coefficients to owner-occupied households to predict imputed rent values.

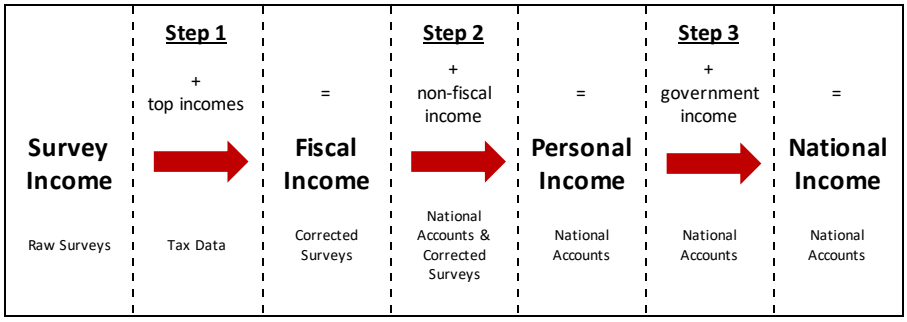
3.4 Estimating DINA: 1976-2016

Figure 3.3 summarises our estimation procedure for DINA into three broad steps. Each step defines the income concept whose distribution we are interested in.

[23] Amadeo et al. (2000); Jaramillo and Saavedra (2005); Gonzalez (2010) provide more information on these programs.

[24] See section 4.3 of Appendix B for details on the formula we use.

Figure 3.3: Framework for Building Pre-Tax DINA for Brazil



Notes: authors' elabouration. Each income concept is associated to an underlying data source for its construction. See text for definitions.

3.4.1 From Survey Income to Fiscal Income

Step 1. In the first step, the transition from "survey income" to "fiscal income" is made by correcting the former for income under-coverage at the top using tax data. To do so we estimate an income distribution from the surveys and the tax data. The former is straightforward, given that we avail of survey microdata. Concerning the latter, we use the generalised Pareto interpolation (gpinter) method of Blanchet et al. (2017) to turn the tabulated distribution into a continuous distribution for the portion of the population included in the tabulation.[25] We then compare income levels in the interpolated tax distribution to those in the survey distribution to verify the extent of income under-coverage in the latter and incorporate the tax incomes into the survey data using a novel combination method developed in chapter 1. Using the distribution of incomes from the tax data, this method reweights survey observations within the original support (to address non-sampling error), and then replaces observations at the top by a distribution drawn from the tax data (to address sampling error), matching the survey covariates to it by preserving the rank of each observation. This method confirms that substantial income under-coverage occurs at the top of the distribution.[26] At the end of the procedure we have a distributional series for total fiscal income – income that would be reported if everyone were to file a declaration – for the overlapping years of both datasets. For years where only one of the two datasets is available we proceed to extrapolate/interpolate according to information from the closest relevant years for which both sources exist.[27]

[25] See `wid.world/gpinter/`.

[26] see Figure B.10 in Appendix B for evidence on selected years.

[27] See Table B.2 in Appendix B for an overview of the strategy we employ for each year where we make these imputations for the estimation procedure of the 1976-2016 period.

3.4.2 From Fiscal Income to Personal Income

Step 2. In the second step, we incorporate into the fiscal income distribution all private non-fiscal incomes that can be attributed to the personal (household) sector, in order to arrive at a distributional series for pre-tax post-replacement personal income.[28] This requires the identification and imputation of missing personal capital income as well as to the imputation of social contributions, which are not included in the gross income assessed for fiscal purposes. The missing personal capital income is income attributed to households but not declared to the tax authorities, and also income that does not get attributed to individuals, but rather to corporations. The first part we can identify as investment income attributable to pension and insurance funds held by individuals and imputed rents, while the latter are the undistributed profits of privately owned corporations associate to Brazilian residents. Thus, in this step we impute the following items:

Social insurance contributions (D61, S14)
Imputed rent for owner-occupiers
Investment income attributable to insurance policyholders (D441, S14)
Investment income payable to pension entitlements (D442, S14)
Household/NPISH component of pre-tax undistributed corporate profits (B5n, S11+S12)

We retrieve the totals of each of these items either directly from national accounts or we estimate them using auxiliary data if the accounts are not sufficiently detailed (this is generally the case for years before 1995).[29]

The value of total social contributions are obtained directly from national accounts publications for 1995-2015 (IBGE 2000; 2017). Pre-1995 totals are estimated using auxiliary national accounts data (IBGE, 2006). Overall, social contributions represent an average 9% of national income over 1976-1994 and 14% over 1995-2015. Totals from imputed rent are obtained from IBGE (2017) for the 2000-2015 period. Before 2000, we use estimates of imputed rent made by Souza (2015; 2016). On average, imputed rent makes up about 7% of national income between 1976 and 1994 and 9% between 1995 and 2015. For investment income on private insurance and pension funds we source the aggregates for 1995-2015 from IBGE (2000; 2017). For previous years we follow two different estimation strategies for each of the two components. For D441, we apply the average share in GDP between 1995 and 2015 (0.08%) to all years between 1976 and 1994. For D442, we make use of data on the total assets managed by private pension funds (*previdencia complementar fechada e aberta*) to estimate D442.[30] The total fund income (D441 + D442) we estimate amounts to about 0.7% of national income over 1976-1994 and 1.2% over 1995-2015.

[28] We include the non-profit sector (NPISH) inside the personal sector.

[29] Full details of the sources and estimation procedure of the totals and their distribution are presented in section 6.2 of Appendix B.

[30] See section 6.2 of Appendix B for details.

The final category in this stage of the estimation process – the household component of pre-tax undistributed profits – is probably the most complex component to measure. In practice, it corresponds to the net primary income of the corporate sector (B5n) attributable to the domestic household sector (S14), through their ownership of equity shares. Thus, we proceed to estimate the separate parts of this component as follows.

First, we obtain total gross primary income of the corporate sector from national accounts for 1995-2015 from IBGE (2000; 2017), and subtract corporate capital depreciation using estimates we make over the period.[31] For earlier years we estimate the net primary income using the growth rate of gross corporate savings and corporate income tax, estimated from (IBGE, 2006), maintaining the same share of corporate capital depreciation in total depreciation as observed for the year 2000. Then, we estimate the personal component of these undistributed profits (as opposed to the government or foreign components), using the household sector share of net corporate equity liabilities (AF5) from the financial accounts for 2010-2015 (IBGE, 2017). For previous years, we make use of proxy data from two business surveys conducted by *Visão* magazine ("Quem e Quem na Economia Brasileira") in 1974 and 1985 on the ownership structure (between private domestic, public and foreign sectors) of the 5,113 and 8,094 largest incorporated firms in the country in each respective year.[32] We proxy the household share of undistributed profits as the private domestic share of the net assets of these firms for the survey years and linearly interpolate the share for intermediate years between 1974, 1985 and 2010.[33] In the end, the household component of undistributed corporate profits amounts to about 5% of national income over 1976-1994 and about 6% for 1995-2015.

At this stage, we have the aggregates for each of the non-fiscal income categories we require in the estimation procedure. Thus, the second ingredient we need is to impute their distribution among Brazilian individuals. We perform this imputation for all four categories using the PNAD survey. Before its correction we have already imputed values for imputed rent per adult and social contributions per adult (see section 3.3.2). After correcting the survey using tax data (section 3.4.1), we impute the aggregates of the other remaining categories – insurance and pension fund investment income and pre-tax undistributed corporate profits associated to households. For fund income, we distribute the aggregated total (D441 + D442) by assuming that it follows the distribution of wages for earners who contribute to a social insurance fund (i.e. who have positive social contributions). For undistributed corporate profits we impute their distribution in line with the joint

[31] See sections 5.2 and 6.2 of Appendix B.

[32] These estimates are reproduced from Bacha (1980) and Baer (2014).

[33] The personal domestic share of undistributed profits is 48% in 1974, 43% in 1985 and an average of 59% over 2010-2015 using these various sources. This is consistent with the privatisation trends in the Brazilian economy from the late-1980s on-wards. Indeed the share of the public sector in total equity assets falls from 48% in 1985 to around 13% in the 2000s.

distribution of financial income (interests and dividends) and employer profit withdrawals.[34] We test seven other imputation "scenarios" for undistributed profits based on variations in the type of income used for the imputation and the estimation of employer capital withdrawals (see sections 6.2 and 6.4 of Appendix B for further details). The different ways of allocating undistributed corporate profits among the population each reflect expectations we make on the 'beneficial (or legal) owners' of them. Our preference is to base our imputation on the remuneration of employers (business owners) that can be attributed to profit withdrawals and the distribution of financial income, which is likely to be the best correlate of owning enterprises. Moreover, it is quite safe to assume that on average, the individuals that take profits out of corporations in some years are in the same part of the distribution as individuals that leave them in corporations. Hence, our benchmark scenario is justified by what we deem to be the most reasonable allocation of undistributed profits, given the variables at our disposal and our knowledge of the tax and social contribution systems.[35]

3.4.3 From Personal Income to National Income

Step 3. Up until this step, we have imputed all the private non-fiscal incomes to the personal sector. The final step involves imputing the remaining categories of income to arrive at a national income distributional series. Here we must account for:

Government factor (capital) income (D4n, S13)
Net production taxes received by the government (D2-D3, S13)
Pension and other social insurance surplus (D61 – (D621+D622), S14)

We assume the distribution of these three categories to be proportional to the distributional of personal income, given the lack of a more precise method to impute each item. Concerning the third category, which has registered a slight deficit in Brazil over the last 20 years, one could argue that the deficit is covered by taxes/levies applied across the entire population on different sources of income/consumption. Likewise for a surplus, which should be attributed proportional to contributing income.

Net production taxes received by the government include value-added taxes and other product taxes, net of product subsidies, (D21-D31) and other taxes on production, net of subsidies (D29-D39). Along with government net property incomes and its operating

[34] The latter are estimated as the portion of employer income that exceeds the exemption threshold for the personal income tax. The survey only reports the total income of employers, which conflates their wage and their profit withdrawals.

[35] This choice is further validated by the sensitivity analysis we depict in section 6.4 of Appendix B. For the other non-fiscal income imputations, we do not test various scenarios for two main reasons. First, multiple scenarios can be less easily constructed given the auxiliary variables we use for them. And second, the assumptions we make on their distribution seem to be a reasonable first approximation, given the incomes we are dealing with.

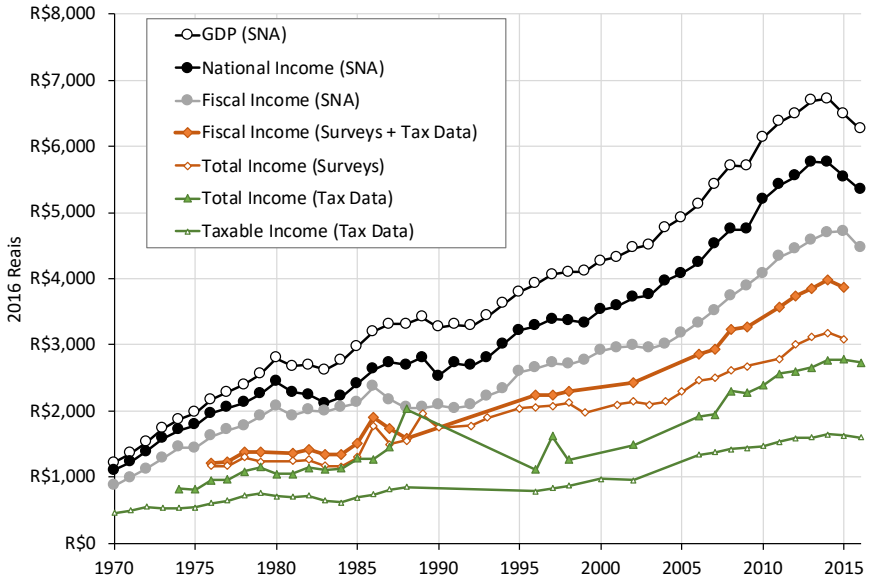
surplus, they form part of government's net primary income. In theory, these taxes should be imputed in relation to their incidence, which can be approximated by consumption, except for property taxes in D29, which should be imputed in proportion to housing wealth or imputed rent (Alvaredo et al., 2017). However, in practice imputations based on tax incidence are difficult to make without access to good data. Moreover, imputing these taxes in proportion to total personal income seams a reasonable first approximation, given that they are based on consumption-type taxes and to production/property-type taxes, which practically cover the entire population. Thus, in proceeding this way for these remaining items, we only seek to normalise the distribution to national income. This step, by definition has no impact on inequality. Figure 3.4 depicts the evolution of the main aggregate income concepts that we mobilise in the estimation process depicted in Figure 3.3 from the survey data, the tax data and the system of national accounts (SNA), as compared to GDP. Through the estimation process just described, we arrive at the distribution of national income, which grew at 2.7% per year on average. It can be seen that other income aggregates grew at broadly comparable rates. Total (pre-tax) income in the surveys represent about 55% of national income, and approximately 70% of an equivalent income total from SNA. Total pre-tax income from tax data, while only covering 20% of the population in surveys, represents close to 50% of national income, and close to 60% of total fiscal income from SNA. Our merged tax+survey fiscal income series covers about 80% of an equivalent total from SNA and equals an average of 65% of total national income over the period. Thus, using this aggregate as a benchmark, approximately 35% of national income is accounted for by "non-fiscal income", which is imputed as previously outlined.

3.5 Estimating DINA: 1926-1975

The previous steps detailed how we construct DINA with all the data we have available for the period 1976-2016. In this section we describe how we proceed for the years prior to 1976, when we have no distributional information from surveys. We are thus left with limited coverage tax data and aggregates from national accounts, alongside the post-1975 estimates.

As described in section 5 of Appendix B, we compute a fiscal income series for top incomes using the tax data and reference totals from national accounts for the entire 1926-2016 period. The population coverage in the tax data allows us to estimate the Top 10% income share back to 1968. Before this we can only continue to estimate the Top 1% share (as well as shares of higher fractiles). In general, our strategy is to anchor our preferred fiscal income series (estimated from tax data and surveys for 1976-2016) and our national income series (estimated from tax data, surveys and remaining income from national accounts for 1976-2016, as detailed in section 3.4), to the evolution of the fiscal income series estimated using tax data and an income denominator from the system of

Figure 3.4: Comparison of Income Aggregates by Income Source in Brazil: 1974–2016



Notes: authors' calculations based on data from surveys, tax tabulations and national accounts (SNA). Nominal values are deflated using the GDP deflator. National income covers about 85% of GDP. The remaining 15% is due to capital depreciation and net foreign income. Fiscal income (SNA), covering an average of 82% of national income, represents an equivalent income total to the income declared on tax returns, which itself accounts for close to 50% of national income. The equivalent income total in the raw surveys covers 55% of national income, while our corrected survey series covers close to 70%. This series merges survey income with assessed incomes from tax data, of which about 60% comprises of taxable income and 40% of non-taxable income.

national accounts (SNA). More specifically, we extrapolate the fiscal income (Tax+Survey) series back to 1926 by adding the average difference in top shares between the Tax+Survey series and the Tax+SNA series for the years 1976-1979, to the Tax+SNA top shares for the preceding years. Similarly, for the national income shares, we extrapolate the 1976-2016 series back to 1926 by adding the average difference in top shares between the fiscal income Tax+Survey series and the national income Tax+Survey+SNA series for the years 1976-1979, to the extrapolated fiscal income Tax+Survey shares for the years 1926-1975. Given that the survey is an integral part of the DINA estimation process there is little else we can do to estimate a distribution of national income prior to the survey's first data point. Thus, we implicitly assume that the average difference between total fiscal income from SNA and the total fiscal income from our merged survey and tax dataset over 1976-1979 is maintained back to 1926, rather than assuming that the "correct" income denominator prior to 1976 is given by our reference total estimate from SNA.[36] The relative stability in the difference between the series over the 1976-1979 period and relatively small average

[36] As eluded to before, the quality/reliability of national accounts deteriorates the further we go back in time.

difference (about -1.0 percentage point for the Top 10% share and +1.8 points for the Top 1%) at least assures us that we are not making questionably large adjustments.[37]

4 Inequality and Growth over the Long-Run in Brazil

4.1 The Ebb and Flow of Income Inequality

In this section, we present our main results and provide some context to them, postponing our overarching interpretation to section 5. Figure 3.5 shows the Top 1% income share in the long-run, while Figure 3.6 shows the distribution of income among four major groups (Top 10%, Top 10-1%, Middle 40% and Bottom 50%), covering the more recent decades with fully-fledged DINA estimates (1976-2016).

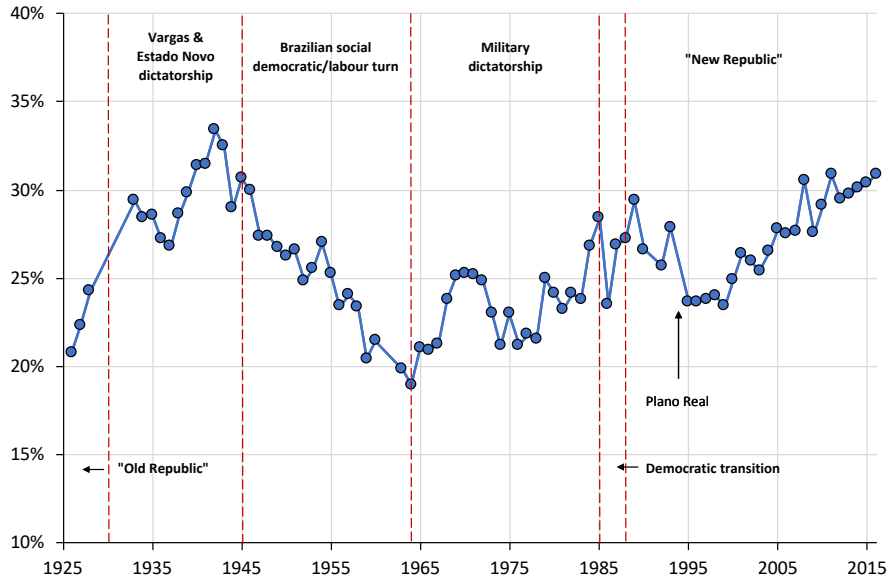
In a nutshell, Brazilian elites have managed to maintain their high relative income share despite decades of tumultuous economic, social, and political change. The distributive conflict at least since the 1970s has mostly been between the middle and the bottom of the distribution. These patterns confirm and also shed new light on earlier results (Medeiros et al., 2015a; Morgan, 2015, 2017; Souza, 2016, 2018).

The Top 1% income share starts at a high and rising level in the waning years of the so-called "Old Republic"(1889–1930), the oligarchic and nominally democratic regime which deposed the Orléans e Bragança monarchy. The Top 1% share climbed from 21% to 24% in the 1920s. Unfortunately, there are no published tabulations for the late 1920s and the early 1930s, when the Old Republic succumbed to the 1930 Revolution, led by Getúlio Vargas. The series resumes in 1933 at a higher level (29%) and receded slightly in the mid-1930s. The concentration of national income at the top then rose steadily during the years of growing authoritarianism, starting in 1937 when Vargas launched a self-coup and established the *Estado Novo* dictatorship, and peaking at 33% in 1942. Hence, unlike in developed countries (see section 4.4), World War II brought no great levelling.

The trend reversed in the closing years of the war, when Vargas pivoted to left-wing populism as he sought to cling to power once the war was over, and the Top 1% income share kept falling in the first couple of years after redemocratisation in 1945. As a result, by the the turn of the 1950s, the concentration of income at the top was back to similar levels as in the mid-1920s. The democratic interregnum of 1945-1964 was, of course, noted for increasing political participation and the consolidation of state-led Import Substitution Industrialisation (ISI) as an overarching development strategy. ISI is often decried as grossly inefficient and inegalitarian, but top income shares declined during its heyday in the 1950s, particularly during the Juscelino Kubitschek presidency (1956-1961). The Top 1% income share reached its nadir of 19% in 1964 – 14 percentage points (p.p.) lower than

[37] See Figure B.13 of Appendix B for a comparison between all these series.

Figure 3.5: The Top 1% Income Share in Brazil: 1926–2016



Notes: Distribution of pre-tax national income among equal-split adults. The unit of observation is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data). Dotted lines mark the time-span of political eras. The *Plano Real* ("Real Plan") was a macroeconomic stabilisation plan implemented in 1994 to tackle inflation. Its main features were the creation of a new currency (the "Real"), which was initially anchored to the US Dollar (complementing the previously existing currency), the complete de-indexation of wage/financial contracts from inflation and the pursuit of orthodox monetary policy.

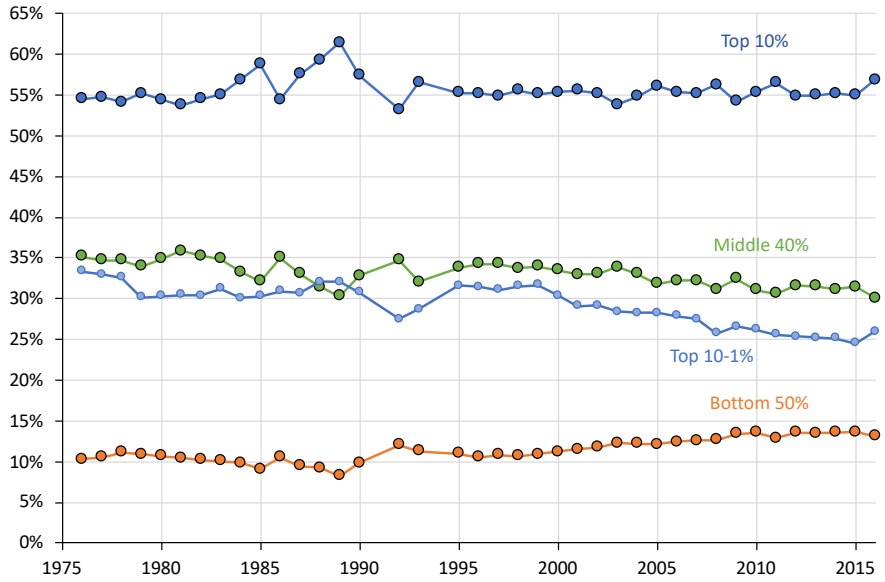
its peak in 1942 and 6-8 p.p. lower than in the early 1950s.

This decline in top shares was interrupted and reversed after 1964, in the wake of the military coup which overthrew the country's first Labour Party executive, under president João Goulart (1961-1964). The Top 1% share rose quickly between 1964-1968 and tapered off at 25% after growth resumed. The authoritarian take on ISI jumpstarted the so-called "Brazilian miracle": from 1968 to 1974, GDP grew at an unprecedented average rate of almost 11% per year. However, mounting foreign debt, rising inflation and increased reliance on oil imports made the situation untenable. Furthermore, by the time the 1973-1974 oil crisis hit Brazil, the newly appointed president, general Ernesto Geisel, was also facing significant political challenges, both from the far right and from the center-left. Geisel decided to take ISI to another level, while at the same time appeasing the masses with more favourable wage policy and social spending. Lower but still significant growth rates were sustained until the 1979 oil shock inaugurated an era of prolonged recession and economic crisis. Thus, the transition back to democracy would happen in a "lost decade" marked by ever-increasing inflation rates and failed stabilisation plans.

The Top 1% income share became more volatile, declining in the mid-1970s and then

rising rapidly in the 1980s as dire economic conditions persisted and inflation spiraled out of control. From 1976 onwards, we have data for the full income distribution. Figure 3.6 shows the income shares accruing to the Bottom 50%, the Middle 40%, the Top 10% and the Top 10-1%, which were more stable than for the Top 1%. The distribution of income between the three groups did not change significantly at least until the mid-1980s, but the first few years after redemocratisation in 1985 were marked by increased concentration of income at the very top: while the top decile did not budge, the Top 1% share soared from 24% in the mid-1980s to 29% in 1989 – this reveals the extent to which the stability of the Top 10% is entirely determined by the rising Top 1% share and the falling Top 10-1% share. The Middle 40% were hit the hardest in relative terms, as their income share dropped from 35% to 30%. The Bottom 50% had a smaller loss in percentage points, with their share dropping from 10% to 8%.

Figure 3.6: National Income Shares in Brazil: 1976–2016



Notes: Distribution of pre-tax national income among equal-split adults. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data).

Macroeconomic stabilisation finally came with the *Plano Real* in 1994. The share of the Top 1% fell from 28%-30% to 23%-25% after the stabilisation, and the Bottom 50% and the Middle 40% recouped their previous losses in the 1980s. Still, the hope that the end of hyperinflation and the new democratic framework established by the 1988 Constitution would usher in a period of declining inequality did not come to pass. After a brief respite, the Top 1% income share recovered quite steadily back to 30% in 2014-2016, about 6 p.p. higher than where it was in 1995. This time, however, the Bottom 50% also improved its lot, as its income share climbed from 10%–11% to 13%–14% between 1995 and 2016.

Consequently, the Middle 40% and the Top 10% excluding the Top 1% were squeezed out. The former had their income share shrink by 4 p.p. to 30% in 2016, but the latter were hit the hardest: increased concentration of income at the very top and stability for the top decile as a whole meant the Top 10%-1% – that is, the "poorest" 90% of the top decile – declined by 5-7 p.p. since the mid-1990s. In any case, in spite of the major recession of 2014-2016, there was substantial income growth in the 2000s, so absolute living standards rose across the board. Interestingly, the sudden reversal of the business cycle did not break trends in income shares, as opposed, say, to what happened in the 1980s.

4.2 The Distribution of Economic Growth

The variation in income inequality, described previously, can be directly associated to the way in which economic growth is distributed. This is because the income share that each income group obtains is dependent on how their real average income moves in relation to the average of the full population. Thus, we can say something meaningful about who really benefited from growth, as officially measured, across different time periods.

As depicted in Figure 3.1, national income per adult has been subject to smooth and quick progressions as well as sharp deviations and stagnation. For the first time, we can reveal how this periodic evolution in overall average income compares to the average of different groups along the distribution. Table 3.1 presents our estimates between 1976 and 2016, the period for which we have distributional data for the entire population

Since 1976, average income has not made much progress. By 2016, the real average income of Brazilians aged 20 or older was 2% lower than what it was in 1976. There are three main reasons for this: the prolonged crisis of the 1980s, the deep recession of 2014-2016, and, to a lesser extent, swift demographic changes in the age composition of the population.[38]

However, not everyone's income was subject to relative stagnation. Consistent with the evolution of income shares, the poorest half of the population made notable real income gains – the strongest out of the three broad income groups. The Middle 40% experienced a decline in their real income of 16%, making them the only group whose income grew less than the national average. While the Top 10% made very modest income gains (reflecting their steady income share), the richest groups within the top of the distribution captured the overwhelming majority of the gains, as the economy transitioned across new economic and political regimes. For instance, the income of the richest 0.1% of the population grew 5 times faster than that of the poorest 50%. This long-run pattern emulates the "squeezed middle" result found for the 2000s by Morgan (2017) (see chapter 2). It also reproduces the global patterns described by Lakner and Milanovic (2013) and Alvaredo et al. (2018).

[38]Growth in the late 1970s was entirely offset by the recession in the 1980s, while the boom of the 2000s has been equally lost to the recent domestic recession. Still, at its peak in 2013 real average income was 13% higher than in 1976. Total adult population grew 178% since 1976, far outpacing total population growth (89%). Total national income increased by 172% in real terms.

What matters for the poor is their relative *levels* of income, rather than its absolute growth. While lower incomes can indeed be subject at times to significant growth, the share of growth absorbed by them depends on this growth being sustained above the growth of other income levels in the distribution for a long-enough period of time. With an average income share of under 12% since 1976 (lower than that of the richest 0.1%), the poorest half of the Brazilian population only managed to capture 15% of the aggregate growth of the whole period, despite notable gains.

Table 3.1: Income Growth and Inequality in Brazil: 1976–2016

Income Groups	Total Cumulated Growth (1976–2016)	Total Cumulated Growth (1976–1988)	Total Cumulated Growth (1988–2000)	Total Cumulated Growth (2000–2016)
Full Population	**-2.1%**	**-3.0%**	**-3.8%**	**4.9%**
Bottom 50%	24.6%	-13.1%	16.9%	22.6%
Middle 40%	-16.4%	-13.2%	2.3%	-5.8%
Top 10%	2.1%	5.6%	-10.3%	7.8%
Incl. Top 1%	*42.8%*	*24.7%*	*-12.0%*	*30.1%*
Incl. Top 0.1%	*127.9%*	*85.6%*	*-14.6%*	*43.8%*
Incl. Top 0.01%	*306.3%*	*207.6%*	*-11.0%*	*48.5%*
Incl. Top 0.001%	*832.9%*	*534.2%*	*-20.3%*	*84.5%*

Notes: Distribution of pre-tax national income among equal-split adults. The unit is the adult individual (20-year-old and over; income of married couples is split into two). The Top 1% is defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal and national accounts data).

Table 3.1 also breaks down the incidence of growth of the period into three sub-periods – the last half of the military dictatorship, the first twelve years of the "New Republic" and the "new" socio-political age of the 2000s. In the first sub-period (1976-1988) gains were only made by richer groups as the Brazilian economy experienced its own brand of extreme stagflation from 1980. Despite the reversals to income growth, top groups benefited from their greater control over remuneration levels and access to indexed-linked income sources, such as financial contracts. The asset-income gains of top income groups more than compensated their losses due to inflation (Kane and Morisset, 1993).[39] For the bottom of the distribution the "inflation tax" operated directly through the channel of real wages, while for the middle it operated through the asset-demand channel, given their lower ability to shift their income across asset-types compared to richer groups (Kane and Morisset, 1993).[40] In contrast, growth in the first decade of the "New Republic", while receding still further, gave way to an inverse distribution – the bottom and the middle

[39]The indexation of interest rates and other financial instruments brought about marked growth in financial intermediation services. These grew from about 3.6% of GDP in the 1950s and 1960s to 7.4% over the 1970s, 13.7% over the 1980s and 21.2% from 1990–1994, according to national accounts data (IBGE, 2006). Such services were previously constrained by rigid ceilings on interest rates (Bacha, 1980).

[40]The middle's "portfolio" mainly consisted of non-indexed contracts (cash and current accounts) and income sources with incomplete indexation (government bonds, social security benefits) (Kane and Morisset, 1993). With the pricing policy of indexation (depending on past inflation, projected

being the only groups to experience real income gains. This inversion is most likely due to the dismantling of indexation (among other policy changes), within the *Plano Real*, which brought inflation rates under control from 1995 on-wards (see Figure 3.1).[41]

At the same time, real structural changes in the economy were also took effect. The Brazilian economy succumbed to the trade and capital-account liberalisation waves of the 1990s, as well as enacting large privatisations of state-owned firms (more on this in section 5). From the late 1980s, industry began to increasingly give way to services in national value-added – the share of services rose from 46% in 1988 to 70% by 1999, which was almost entirely compensated by the mirrored fall of industry – itself driven by the halving of the manufacturing share in GDP from 30% to 15% (Figure 3.2). Relative employment shares changed less over the period, with services continuing to be the largest employing sector, accounting for over 50% of employment.[42] This shift was facilitated by the combination of trade liberalisation and market de-regulatory processes in the late-1980s and early-1990s, with the high-interest-rate/credit-restrictive policy of the mid-1990s. While this reduced the monopoly power of certain domestic firms (and their previous profitable mark-up policies), it also overly appreciated the currency and dampened investment expectations by the latter 1990s. Thus, the compression of the overall income distribution is likely due to the combination of lower bargaining of previously powerful urban industrial workers (manifested in reduced skill-experience wage premiums) and the increase in the share of self-employment in commercial activities and services on the one hand, with the monetary changes of the *Plano Real* on the other (Saad-Filho and Mollo, 2002).

The re-emergence of growth in the 2000s brought positive gains back to most income groups. However, the inverse symmetry between the dynamics at top of the distribution and the rest, which defined the previous two decades, no longer holds. The distribution of growth has increasingly become symmetrical between the tails. On the one side, new economic conditions (lower-skill sectoral growth) and political conditions (arrival of a "pro-poor" government) were captured in the strong growth of bottom incomes, while on the other side, the rich experienced substantial income gains. The latter were likely due to lagged effects and continuation of pro-rich policies and outcomes from the 1990s – capital account liberalisation, high interest rates, tax reform, privatisations, trade liberalisation and industrial concentration (Baer, 2014) – as well as to the primary commodity boom; the heavy subsidisation of industrial credit by the government in a renewed search for

future inflation and productivity), past inflation was project into the future and was exacerbated by negative supply shocks. The rapid inertial component of inflation can be seen in the narrowing of the wage-adjustment window over time. Until 1979, wages were adjusted once per year, after which they were subject to two annual adjustments until 1985, then adjustments occurred approximately every 3 months until 1987. Monthly adjustments subsequently took place after 1987 (Saad-Filho and Mollo, 2002).

[41] Indexation was ratified, however, for social security transfers by the new 1988 Constitution: all benefits are adjusted annually to recoup losses due to inflation and no benefit can be lower than the minimum wage, which progressed in real terms from 1995 on-wards.

[42] These structural changes suggest that income effects dominated employment effects.

"national champions"; and emerging asset bubbles in the stock and real-estate markets. The losers during the process were the "squeezed" Middle 40% and the upper middle class comprising the Top 10-1%. These changes help to explain the reduction in inequality among the Bottom 90% that can be observed from the mid-1990s in Figure 3.6.[43]

While our data only allows for an assessment of the complete distributional impacts of growth from the mid-1970, we can assess the distribution of growth over a longer time-frame for a reduced subset of the population.[44] In Table 3.2, we present the annual average growth of national income for the full population and for the Top 1% – a group sufficiently large and affluent to be politically and economically relevant – across the full span of our data. We additionally divide up the full period into five sub-periods, broadly relating to economic and political changes. In the very long-run, economic elites (comprised in the richest 1%) have managed to out-grow the "average" Brazilian by 0.4 percentage points per year. Yet, this evolution has been far from linear, as one would expect from Figure 3.5. For instance, average elite income rapidly out-grew overall average income income as the country was slowly transitioning from an old agrarian-based society to a more mixed-sector economy (1926-1945) – when emerging urban-elites had not quite displaced the old-rural elites.

Table 3.2: Income Growth and Inequality in Brazil: 1926–2016

Period	Annual Average Income Growth	
	Full Population	Top 1%
1926–2016	2.0%	2.4%
1926–1945	2.4%	4.4%
1945–1964	3.5%	1.0%
1964–1985	2.6%	4.5%
1985–2000	-0.1%	-0.9%
2000–2016	0.3%	1.6%

Notes: Distribution of pre-tax national income among equal-split adults. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal and national accounts data).

Greater economic expansion would follow over the next 35 years, but under very distinct political economies. Between 1945 and 1964, during the "urban-worker" coalition, total average annual income growth was 3.5%, while it was 1.0% for the Top 1%. Thus, growth was engineered more in favour of lower income groups (at least among the Bottom 99%). The military government (1964-1985) would completely reverse the preceding trend, allowing strong growth to be translated into larger income gains for the rich (6% per year) than the average Brazilian (3% per year). In the last 40 years, the overall mass of the

[43] For more details on this period see chapter 2.

[44] This subset is anchored to the share of the population present in tax data.

population either receded slightly less than elites during periods of stagnation (1985-2000) or grew slower during periods of renewed growth (2000-2016). Thus, the distribution of growth seems to closely relate to the political cycle and its interaction with economic conditions. In section 5.2 we offer a more detailed interpretation and analysis of the institutional determinants and conventional justifications of Brazilian inequality trends, in light of our results.

4.3 Gender Inequality

So far we have focused on income inequality among adults across the national territory, disregarding sub-populations. In this section we introduce new estimates on an important dimension of total inequality: the inequality between men and women in labour earnings. In our previous calculations, we split income equally between spouses. However, to best study gender inequality, we assign to each individual their own labour income. We concentrate on labour income because the structure of income tax data in Brazil is such that only labour incomes can be approximated for the two sexes.[45]

To divide labour earnings by gender, we use information on the composition of total "taxable income" from the tax data, available for the years 1974–1988, 1998 and 2014–2016. As explained, this income concept mostly comprises of pre-tax labour income (wages of salaried and self-employed workers and pensions) and property rent, at least from the 1990s. In the earlier period taxable income also included other capital incomes like interests and profits/dividends. Thus, to arrive at estimates for pure factor labour remuneration for the working-age population, we must exclude capital incomes and pensions. To do so, we apply ratios of factor labour income estimates to "taxable income" estimates from the surveys to our taxable income estimates from tax data for years when the latter comprised of labour incomes (including pensions) and rent (1998-2016). When the taxable income concept included more types of capital income (1974–1988), we used estimates from Morgan (2015) on the composition of top incomes to reduce them to labour income (including pensions) and rent, to which we applied the ratios from the surveys.[46]

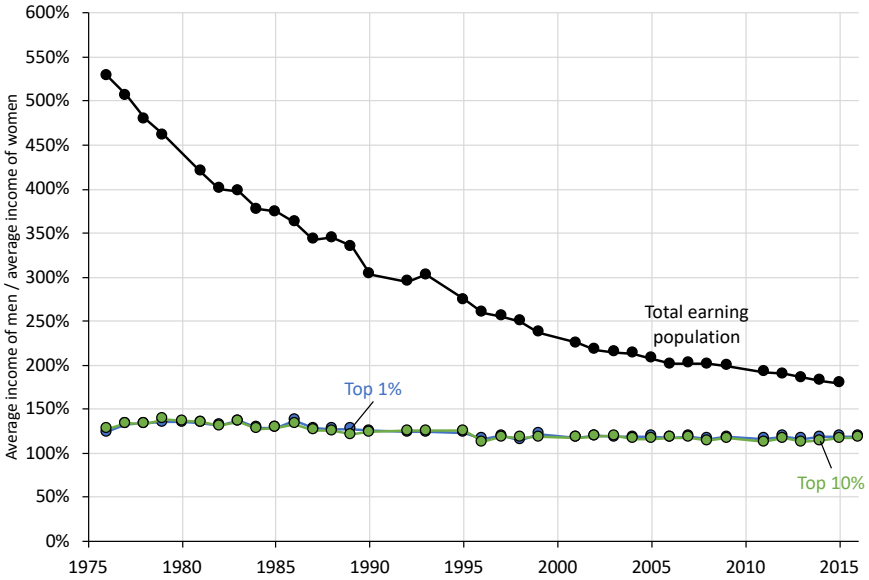
Figure 3.7 quantifies the extent of the gender gap in labour earnings, and its change over time, in Brazil across the full population and higher-income fractiles. This statistic divides the pre-tax average labour income of working-age (defined 20-64) men by the average labour income of working-age women. This measure of the gender gap is broad yet relevant, in the sense that we focus on differences in monetary labour remuneration from wages/salaried work and self-employment income, conditional on working hours and labour force participation. In Brazil during the 1970s, men earned between 4.5 and 5.5 times more than women on average. In 2015, men earned 1.79 times more. The gender gap has more

[45] In general, capital income is more difficult to attribute individually given the lack of information on the division of property between married couples.

[46] See Appendix B for details.

than halved over the forty year period. Interestingly, among top income groups, the gender pay gap has been substantially lower, with the decline being much less dramatic. In 1980, men in the Top 10% and Top 1% made 1.37 times and 1.35 times more than women, while in 2015, they made 1.17 and 1.19 times more. Women in Brazil seem to fare better at the top of the distribution than among the total earning population on average.

Figure 3.7: Gender Gap by Fractile of Labour Income in Brazil: 1976–2016



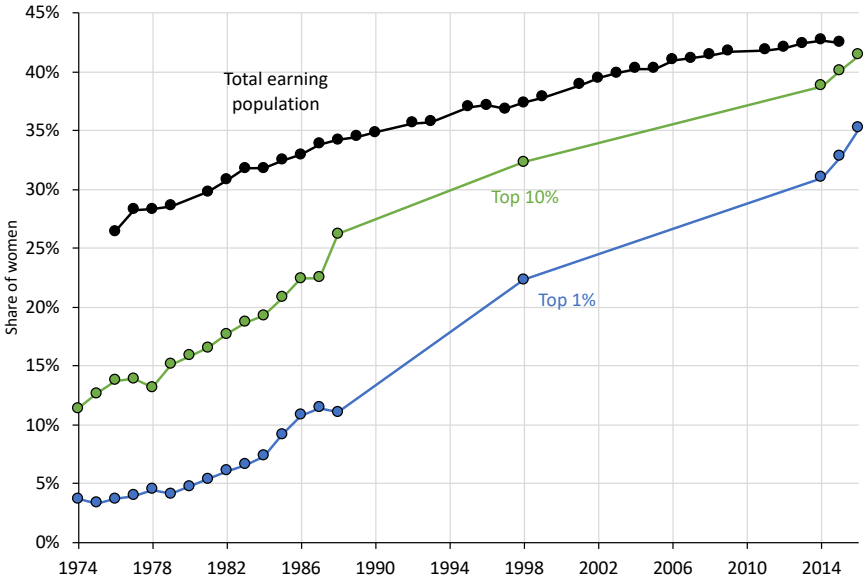
Notes: The gender gap is the average labour income of men divided by the average labour income of women. Distribution of factor labour income (wages/salaries, labour component of mixed income). Fractiles are defined relative to the total number of working-age individuals in the population (between 20 and 64 years of age). Authors' calculations (combining surveys and tax data).

Gender inequalities still persist at the top of the distribution. But again, the change over time has been remarkable. As Figure 3.8 shows, the share of women in the work force (i.e. with positive labour income from salaried work or self-employment) has grown from 27% in 1976 to 42% in 2015. The overall participation gap has almost been closed. These results are robust across data sources: Census data from 1970 to 2010 reveal the same pattern.[47]

Unsurprisingly, women are less represented in top labour income groups, with participation falling for higher groups within the top. In the 1970s women accounted for less than 15% of the Top 10% labour income earners. By 2016, they accounted for 44% (+ 33 points). Women in the Top 1% made similar gains, with their share rise to 35% in 2016 from 5% in the 1970s (+ 30 points). This means that the representation of women in the richest 1% of labour income recipients has grown twice as much as the representation of

[47]The female share of the labour force – including unpaid workers – rises from 17% in the 1970 Census to 42% in 2010.

Figure 3.8: Share of Women in the Employed Population by Fractile of Labour Income in Brazil: 1976–2016



Notes: The employed population covers individuals between 20 and 64 years of age with positive factor labour income (wages/salaries, labour component of mixed income). Fractiles are defined relative to the total number of employed individuals in the population. Authors' calculations (combining surveys and tax data).

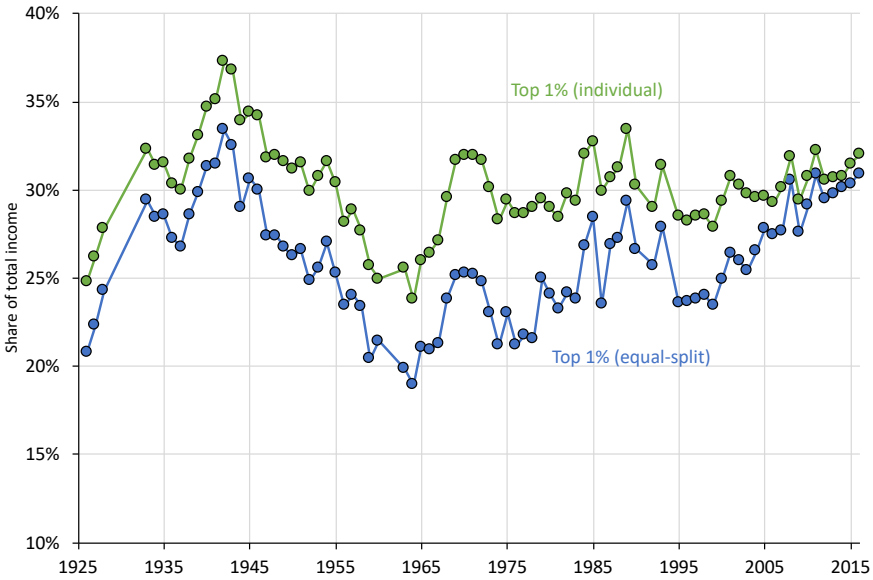
women in the total earning population. Again, comparison with Census data yields similar results.[48]

The significant reduction in gender inequality in Brazil over the past 40 years can further be appreciated from estimates of total income concentration when each spouse is assigned their own income. Figure 3.9 present this case in comparison to our benchmark equal-split estimate for the Top 1% income share. The individualised series is computed using the same procedure as the equal-split series (described in sections 3.4 and 3.5), except that all individuals are assigned their original income.[49] Thus, non-working spouses are attributed zero labour income. Comparing the case where spouses' income is split equally to the case where they are assigned their own income makes it possible to examine how changes in the labour participation of women and gender inequality impact overall income inequality. Figure 3.9 confirms that gender inequality and female labour force participation makes a notable difference in equalising levels and trends over time. In Brazil,

[48]The female share of the Top 10% was just 11% in the 1970 Census, but climbed steadily to 33% in 2010. In the case of the Top 1%, it increased from 3% to 25%. Due to data limitations, these numbers refer to the distribution of total individual incomes, not labour incomes.

[49]This straightforward to compute from the survey microdata. Given the nature of the tax tabulations, however, we assume for the individualised case that all jointly-filed declarations correspond to one income recipient, which tends to be men. See section 4.2 of Appendix B for details.

Figure 3.9: Top 1% Income Share: Individuals vs. Equal-Split



Notes: Distribution of pre-tax national income among adults. The equal-split series divides the total income of married couples equally. The individual series assigns each spouse his or her own income. Fractiles are defined relative to the total number of adults aged 20 and over in the population. Authors' calculations (combining survey, tax and national accounts data).

as elsewhere[50], inequality is higher when income is attributed on an individualised basis rather than being equally split between spouses. During the 1970s the Top 1% share was 7 point higher on average in the individualised series than in the equal split series. In the 1980s this had fallen to 5 points. And since 2010, the difference has been 1 point. The strong decline in the gap reflects the fall in gender inequality in general, but particularly among the top percentile, as shown in Figure 3.8. Thus, on the basis of this evidence, the decline in gender inequality has been an important factor in attenuating the rise of inequality in Brazil since the 1970s. Nonetheless, the rise of female top earners – more than likely married to other top earners – has meant that a greater rise in concentration is observed when equally-splitting the income of spouses.

In terms of understanding the above findings, there are related socio-demographic changes we may highlight. Over the entire period, women increased their participation in the formal labour force from 25% to 44%, slightly outpacing their participation in the total labour force, which includes the lower-wage informal sector. There is also evidence, from the surveys, that women increased their share in public sector jobs (from 54% in 1992 to 58% in 2015).[51] Thus, we can interpret that women have become more represented in general in high-paying occupations. At the same time, women have also become more educated, as average years of schooling for the entire population increased. In particular,

[50] See Piketty et al. (2018).

[51] Unfortunately, the surveys only distinguish public sector employment from 1992.

while only 32% of those employed with higher education degrees in 1976 were women, in 2015 the female share was 55%. This is in a general context where, on average, women were always more educated than men (in terms of average years of schooling, at least since the 1970s).

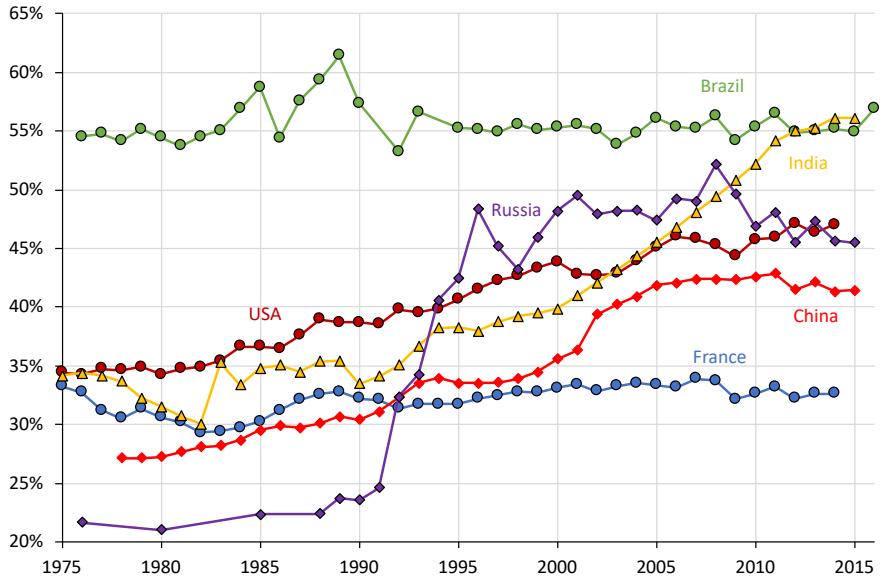
So, it is not so much that the educational gap is lower today, but rather that female participation rates have increased substantially: there are now, for the first time, more employed women with higher education than men. Even if there was a slight reduction in the gender-income gap by education (women still earn about half of what men earn among the higher educated – reflected in the gender gap for top fractiles – rather than 35-40%, as was the case during the 1970s), compositional effects drove the changes.[52] We would argue that societal gender roles have much to do with these dynamics: men had to start earning a living as early as possible (more often than not in manual jobs), whereas women were more likely to be housewives, nurses or school teachers, so middle-class families had incentives to let them study longer. Therefore, "female-oriented jobs" typically required more formal education, and were associated to sectors that would experience massive growth from the 1980s (i.e. services).

Also, the prevalence of relatively cheap domestic service in upper-income households in Brazil can be a facilitating factor to both spouses' career aspirations. This would apply to married/co-habitating couples, but also to single working women. In fact, women in the former group have become more important over time as can be calculated from the PNAD surveys. In the 1970s, about 47% of women in the Top 10% of earners were married/co-habitating. By the 2000s, this share had increased to 57%. A similar evolution is observed for the Top 1% of earners, with about 50% of women being married/co-habitating in the 1970s, and 60% in the 2000s.[53] This shows that the participation of single working women in top earners is still significant, which is consistent with women's average level of educational attainment and their increased participation in the labour market.

[52]The gender gaps are lower the further one moves down the educational distribution. Today, women with secondary education earn two-thirds of what men earn, while those with primary education or less earn three-quarters of what men earn on average. The changes over time are similar to those in the higher educational group.

[53]The tax tabulations present information on civil status over the 1970s and 1980s for taxable income. It too shows an increase in the share of married women at the top. But the main difference with surveys is that, given the wider income concept, widows are more represented at the top in tax data. In the late 1970s the share of widowed women out of all women in the Top 1% was almost 30%. This can be explained by the inclusion of pensions as well as property income. By contrast widows can only be identified in survey data from 2009. While their share of women in the Top 1% for taxable income was about 15%, it was just 2% for labour incomes.

Figure 3.10: Top 10% Shares in Brazil vs Selected Countries: 1975–2016



Notes: Distribution of pre-tax national income among equal-split adults. The unit of observation is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations for Brazil (combining survey, tax and national accounts data). Data for other countries is from https://wid.world/.

4.4 International Comparisons: A Missed Levelling?

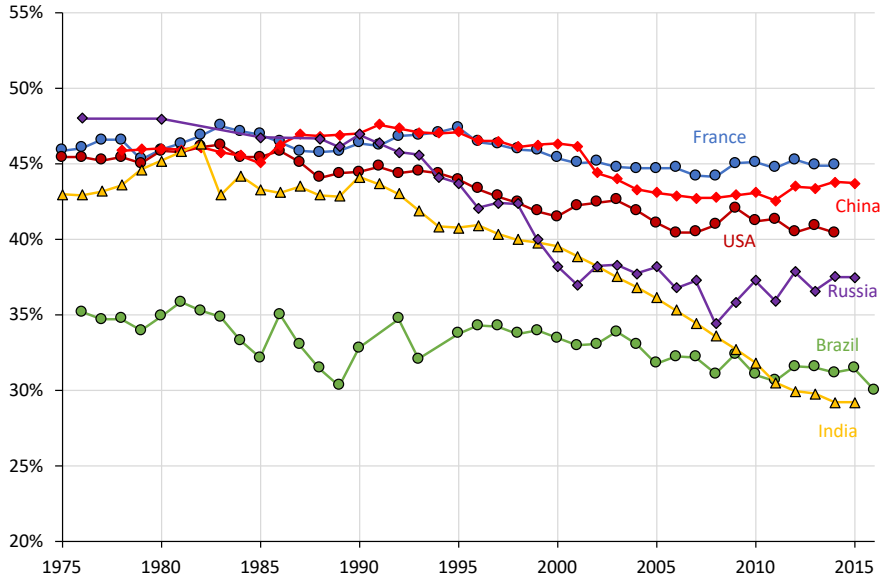
4.4.1 Income Inequality

Cross-country comparisons of income inequality based on survey and/or fiscal estimates are hugely important, but may be biased by time- and country-specific legislation and survey sampling and non-sampling errors. DINA estimates are of great help to comparative research because they rely on an encompassing income concept – net national income – that can be rendered more homogeneous across countries and time.

Figures 3.10, 3.11 and 3.12 provide comparisons across the full distribution of income for Brazil and five very diverse countries – two fast-growing emerging economies (China and India), one former communist country (Russia) and two developed countries corresponding to different "varieties of capitalism" (France and the USA).

Figure 3.10 presents the series for the top of the income distribution. Brazil stands out, as it is the only country where the Top 10% income share is relatively stable at a very high level since the mid-1970s. Elsewhere, the concentration of income was at least some 20 p.p. lower at the start of the period and then either remained stable (France) or increased rapidly (all other countries). Consequently, the difference between the Top 10% share in Brazil and the average Top 10% share in the other five countries dropped from 26 p.p. to 10 p.p. between 1980 and 2014.

Figure 3.11: Middle 40% Shares in Brazil vs Selected Countries: 1975–2016



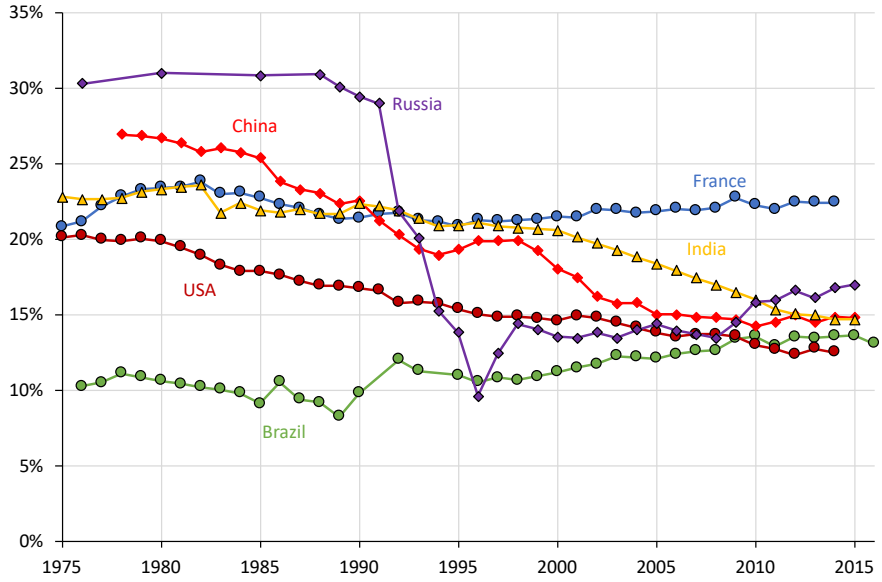
Notes: Distribution of pre-tax national income among equal-split adults. The unit of observation is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations for Brazil (combining survey, tax and national accounts data). Data for other countries is from https://wid.world/.

Results for the Middle 40% (Figure 3.11) show declining income shares everywhere, except for France, which remained relatively stable. The fall was steep in India and Russia (10-15 p.p.) and milder in Brazil, China and the USA (3-5 p.p.). Again, Brazil was more of an outlier in the mid-1970s than in the 2010s, and for the same reasons: the other countries converged to Brazilian levels, rather than the other way around.

Figure 3.12, however, presents somewhat differing trends. While the poorest in Brazil increased their income share by almost 5 p.p., they lost ground in four of the five other countries – France, again, is the exception. Brazil totally ceased to be an oddity. The Bottom 50% now have essentially the same share of total income in Brazil as in most countries in our sample.

Four stylised facts emerge from these comparisons. First, DINA estimates confirm that income inequality in Brazil is very high by international standards, which in turn stems from the large concentration of income at the top. Second, income shares followed a unique trajectory in Brazil over the past 40 years. After all, there was relative stability across the board in France, and the Top 10% squeezed out both the Middle 40% and the Bottom 50% elsewhere. Brazil was the only country where the share of the Top 10% hardly changed and the distributional conflict pitted the Bottom 50% against the Middle 40%. Third, these contrasting trends mean most countries in the sample converged (to a certain extent)

Figure 3.12: Bottom 50% Shares in Brazil vs Selected Countries: 1975–2016

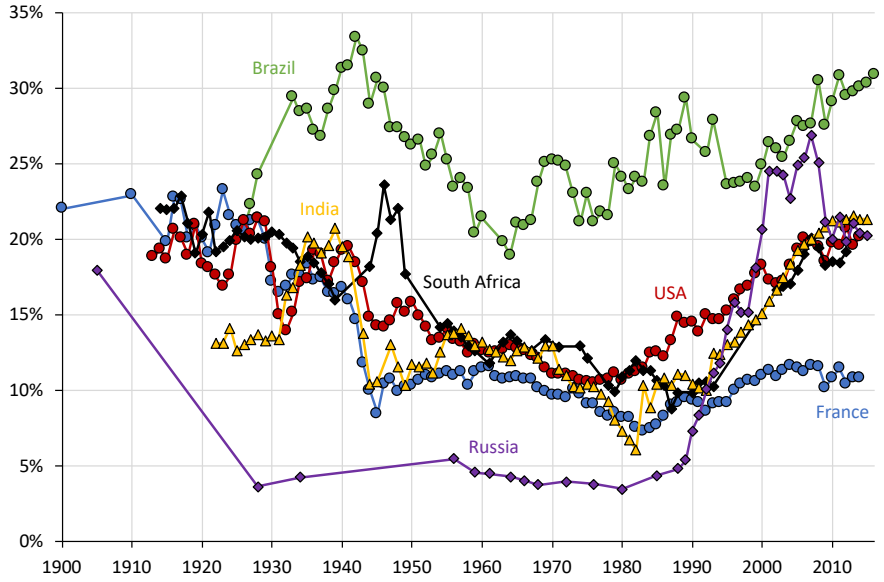


Notes: Distribution of pre-tax national income among equal-split adults. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations for Brazil (combining survey, tax and national accounts data). Data for other countries is from https://wid.world/.

to Brazilian levels of inequality. Although income is still more concentrated at the top in Brazil, the difference now is not as stark as it was in the mid-1970s. Fourth, while all non-developed countries in the sample underwent great institutional change during this period, Brazil was the only country where this did not translate into changes in the Top 10% share. Thus, greater concentration of income at the top followed both China's and India's market-oriented reforms, and even more so in Russia, where the fall of Communism led to soaring inequality. On the other hand, in Brazil, the redemocratisation of society, after a right-wing military dictatorship, and a new social-democratic Constitution, did not entail any significant decline in inequality. Though this can be largely attributed to the political compromise underpinning the "New Republic", it provides some evidence that inequality at the top is asymmetrical: whenever Scheidel's (2017) "four horsemen" – mass-mobilisation warfare, revolution, state collapse and pandemics – are not in action, greater concentration of income at the top seems to be far more common than its opposite.

As mentioned previously, we could only calculate a very long-run series for the Top 1% in Brazil. Figure 3.13 compares our estimates with almost the same five countries as before (with South Africa replacing China), for the period since the start of the 20^{th} century. The longer time span and the focus on the very top of the income distribution allow us to refine some of the previous conclusions.

Figure 3.13: Top 1% National Income Shares in Brazil vs Selected Countries: 1900–2016



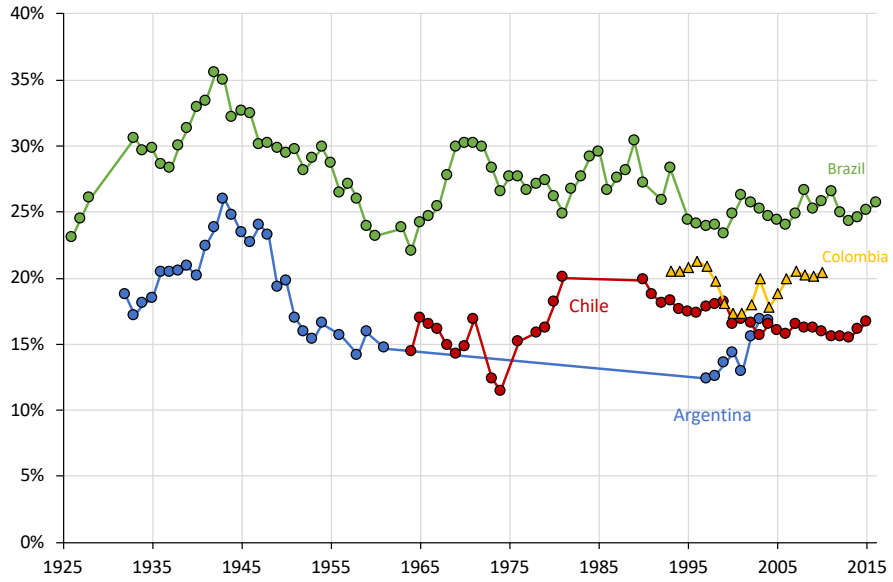
Notes: Distribution of pre-tax national income among adults (the Indian and South African series are of fiscal income, as defined by their tax authorities, scaled up to national income). The unit of observation is the adult individual (20-year-old and over; income of married couples is split into two for all countries except South Africa, which only accounts for individual income). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations for Brazil (combining survey, tax and national accounts data). Data for other countries is from `https://wid.world/`.

The concentration of income at the very top in Brazil is unquestionably higher than in other countries, much more so than it appeared even in Figure 3.10 – that is, inequality among the rich seems to be higher in Brazil, as the Top 1% typically get a higher share of top decile's total income. Therefore, even if the other countries moved closer to Brazil regarding the income shares of the Top 10%, Middle 40% and Bottom 50%, the Brazilian Top 1% series sits far above the others for most of the period – except, for the late 1920s and early 1930s, when the USA, France, South Africa were quite close to Brazil, and for the early 2000s, during Russia's peak inequality levels. The longer time span confirms that developing countries do not adhere neatly to neither the U-shaped pattern seen in the USA nor to the L-shaped trajectory followed by France. Brazil is no exception. If anything, India, Russia and South Africa seem closer to the U-shape evolution than to Brazil's wave-shaped pattern.

Research on top incomes typically points out that most of today's developed countries were very unequal until the eve of World War I but the turmoil between 1915 and 1945 – especially during World War II – caused a plunge in top shares. Recently, Williamson (2015) posited that contemporary Latin American inequality is higher than in most of the world because the region missed the "great leveling" of the post-war period. Thus, for

Williamson, the role of the colonial legacy in sustaining extreme inequality must be played down. Results from Figure 3.13 do show that, apart from South Africa, all other countries experienced abrupt dips in the Top 1% income share in the early 20^{th} century, particularly amid World War II, while the Top 1% in Brazil surged. However, it is not accurate to say Brazil entirely missed the great levelling. Rather, Brazil had its own belated levelling, as the income share of the Top 1% dropped by more than 10 p.p. between 1945 and the early 1960s. The point is that this belated levelling was also short-lived, insofar as it was almost fully reversed after the 1964 military *coup*.

Figure 3.14: Top 1% Fiscal Income Shares in Latin America: 1926–2016



Notes: Distribution of pre-tax fiscal income among individuals (spouses are attributed their own income). The unit of observation is the adult individual (20-year-old and over). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations for Brazil (combining survey, tax and national accounts data). Data for other countries is from `https://wid.world/`.

Lastly, Figure 3.14 displays the Top 1% series for in Brazil and in three Latin American countries – Argentina, Chile and Colombia. Due to the nature of data in other countries we only compare estimates for fiscal income shares at the individual level (without equally-splitting the income of spouses). Given the relative data constraints – e.g. only Argentina has a long-run series, but no national income figures – a case by case comparison is more insightful than any attempt to identify common patterns. As such, the first striking result is how Brazilian and Argentinian top shares moved in tandem until the mid-to-late 1950s. Both countries had similar political regimes and faced similar external conditions, resulting in similar spikes. However, the changes came from different sources (agriculture in Argentina, mostly manufacturing in Brazil).[54] In sum, though Brazil has always been

[54]See section 4.1 and Alvaredo (2010).

more unequal, the magnitude of the changes in top shares before and after the war was almost the same. Indeed, the resemblance between key political movements in this period, associated mainly to initiatives by Getúlio Vargas and Juan Domingo Perón, illustrates how the two countries moved alike. Unfortunately, there is no data for Argentina for most of the second half of the past century so it is hard to delimit when the two countries parted ways.

A second point of interest is that the increase in inequality during the Augusto Pinochet dictatorship in Chile (1973-1990) was comparable to the rise of the Top 1% in Brazil in the first decade after the 1964 military coup. Nevertheless, since redemocratisation Chile seemed to have been slightly more successful in reducing top shares, particularly after 1995. Since 2013, top shares in both countries have followed a similar upward trajectory.

Putting all the pieces together, it is hard to place developing countries under the broad categories used to describe the evolution of top incomes in the rich world. And Brazil is no exception. Even after accounting for gaps and other issues with the data, the country deviates from the others in both levels and trends. The concentration of income at the very top in Brazil is extreme and persistent, though not completely stable, displaying a unique wave-shaped evolution. The major similarities are with Argentina in the 1920s-1950s and Chile during the Pinochet dictatorship. For the rest of the distribution, the convergence is greatest for the share of the Bottom 50% in recent years, with other countries converging downwards and Brazil converging upwards. Therefore, from the perspective of the top, it can be said that Brazil had a short-lived levelling, but from the perspective of the bottom, there is evidence for a somewhat late levelling in inequality relative to the rest of the population.

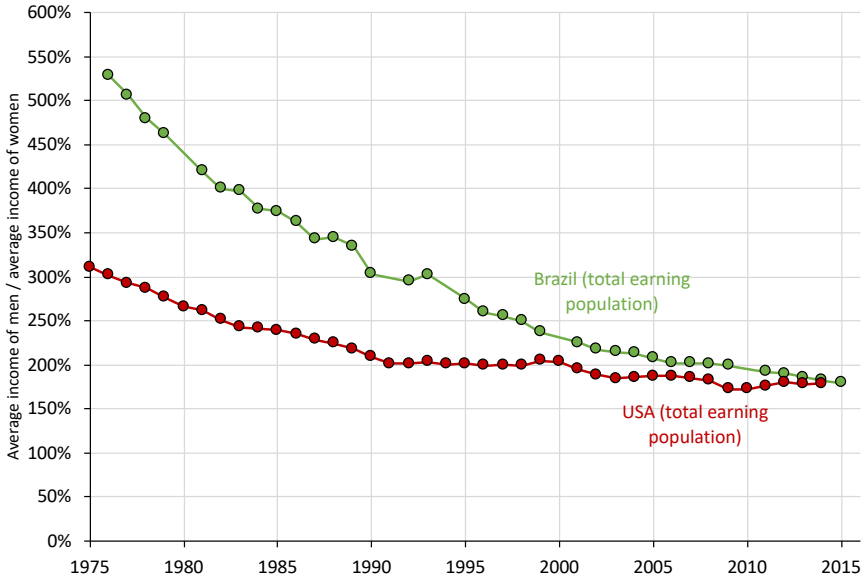
4.4.2 Gender Inequality

In the previous section we documented how Brazil is somewhat of an outlier regarding the long-run evolution of inequality across countries. Since the 1970s, levelling tendencies have only really come from the steady rise in the Bottom 50% income share, especially over the 2000s. In section 4.3, we showed that notable progress was made regarding gender-pay gaps and the representation of women in the distribution. In this section, we ask how the levels and evolution in gender inequality compares to other countries with comparable data.

Figure 3.15 displays the difference in the gender gaps for the working population (aged 20-64) in Brazil and the USA. In the 1970s, men earned about 3 times more labour income than women in the USA. In Brazil the difference was in the order of 5. By 2014/2015, men earned 1.8 times more in both countries. Thus, Brazil has managed to reduced the pay gap between men and women much quicker than in the United States – Brazil achieved the same reduction in the gender gap (from 3 to 1.8 times) in almost two-thirds of the time.

Even more striking are the comparisons in female representation in top labour income groups. Here we avail of an additional case – France – to complement the USA. Focusing

Figure 3.15: Gender Gap in Brazil vs USA: 1975–2016

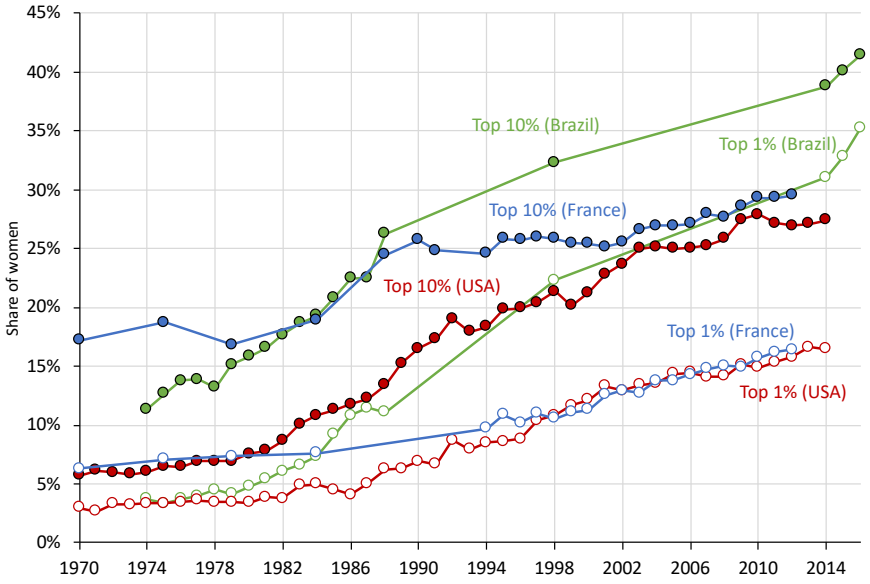


Notes: The gender gap is the average labour income of men divided by the average labour income of women. Distribution of factor labour income (wages/salaries, labour component of mixed income). Fractiles are defined relative to the total number of working-age individuals in the population (between 20 and 64 years of age). Authors' calculations (combining surveys and tax data). Data for USA is from Piketty et al. (2018).

on the Top 10% and the Top 1% in the labour income distribution, we find that Brazil stands out in relation to the two developed cases. In the mid-1970s, women in Brazil were more represented in the top decile than in the USA, but less than in France. By 2014, Brazilian women represented almost 40% of the richest 10% labour earners, compared to about 30% in France and 27% in the USA. An even greater disparity has emerged in the share of women in the Top 1% in Brazil versus developed countries. In the 1970s, Brazil was no different than the United States, both slightly lagging behind France (3% vs 7%). By 2016, the representation of women in the top percentile had surged to 35%, advancing by 6% per year. By contrast, the female shares in the richest 1% France and USA were about half of the share of their Brazilian counterparts, growing at roughly half the speed (in the case of France).

As mentioned in Section 4.3, such results are not a quirk of the particular data set we used – Census results from 1970 to 2010 essentially replicate the same findings, suggesting the share of women at the top of the income distribution is indeed higher in Brazil than in the USA and France (see section 9.1 of Appendix B). The same is true of annual household survey data. This mitigates the possibility that the results we find from tax data for Brazil are due to artefacts of this data source – in particular to the option of married couples to pool or split their incomes in such a way that they optimise their household's total tax

Figure 3.16: Share of Women at the Top of the Labour Income Distribution in Brazil, USA and France: 1970–2016



Notes: In Brazil and the USA, the employed population covers individuals between 20 and 64 years of age with positive factor labour income (wages/salaries, labour component of mixed income). Fractiles are defined relative to the total number of employed individuals in the population (20-64 for Brazil and the USA and 25-59 for France). Authors' calculations (combining surveys and tax data). Data for USA and France is from Piketty et al. (2018) and Garbinti et al. (2018), respectively.

liability.[55] Thus, overall, Brazil has achieved a late, but significant, levelling in gender gaps, at an even greater speed than developed countries for which we have comparable data. Investigating the precise reasons for these differences is beyond the scope of this paper, and a subject for future research. One can think that it could have a lot to do with the differences in gender roles held by society and their evolution over time across countries. And these can be impacted by socio-economic factors linked to levels of development. For instance, the difference in levels between female representation at the top of the distribution Brazil versus the USA and France could relate to differences in the role of domestic service, in marriage rates, in gender differences in life expectancy, in female education levels, etc.

[55]With the option of joint filling, some declarations made by married women could in theory include the income of their spouses. But this is less likely than the inverse – of women appearing on the declaration of their male spouses. Moreover with joint declarations representing between 20-30% of total declarations in recent years, the impact of this factor would not be large if it exists. The other option, of husbands and wives splitting their labour income to optimise on tax payments, would only be possible in the case of both working in the same firm. See section 9.1 of Appendix B for further discussion on these concerns.

5 Discussion

5.1 Methodological Perspectives

As discussed in section 3.1, our DINA estimates combine information from multiple sources to assess the distribution of a broad income concept, namely, net national income from national accounts. The theoretical link between this concept and the less encompassing income concepts traditionally used in research on income inequality is straightforward (see Figure 3.3). However, there are substantive and methodological reasons to also examine the empirical consequences – both in terms of levels and trends – of using different concepts. In particular, it is worth comparing net national income figures with estimates based on the two most common definitions in the income inequality literature – fiscal income and survey income. The comparison with the latter is especially important, given that surveys remain the most popular data source among social scientists.

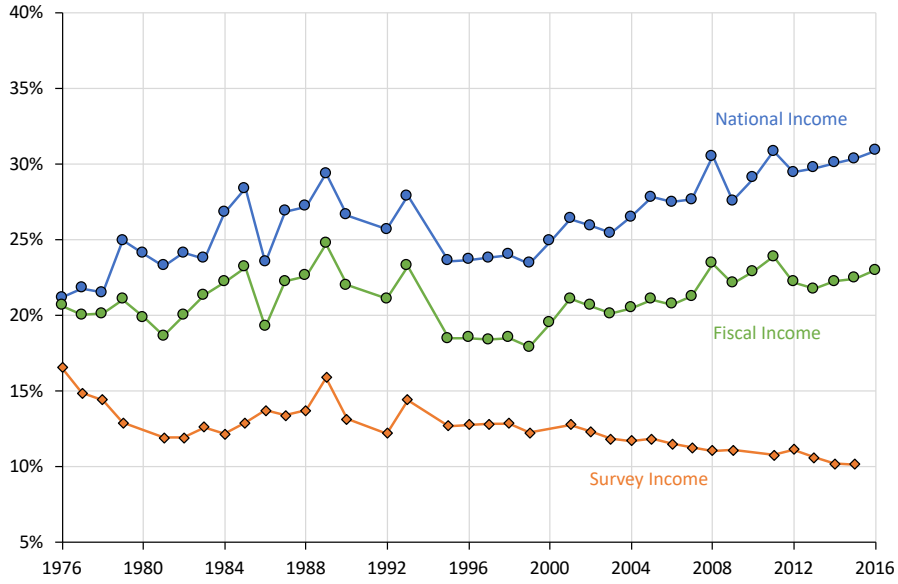
Let us start at the very top. Figure 3.17 shows the Top 1% income share for the three different income concepts from 1976 to 2016. At first glance, two patterns stand out. First, national and fiscal income estimates are more similar in trends than in levels, although their level-difference is due to an early discrepancy between them since the late 1970s. The gap between the two estimates for the Top 1% income share went from 2-3 p.p. in 1976–1985 to 7-8 p.p. in 2006–2016. This could be partly attributed to higher quality National Accounts data in recent years, but it also seems to reflect two real phenomena: the rise in Social Security contributions and undistributed profits of domestic companies as a percentage of national income.[56] For the time being, at least, the fiscal income series is not a poor predictor of changes in the Top 1% income share, but it underestimates its level.

Second, surveys grossly underestimate both levels and trends, and this appears to be getting worse over time. The figures for the late 1970s were already lower than in the other two series, but at least they were much closer than they are now: while in 1976 the national income estimates for the Top 1% were 5 p.p. higher, in 2015 the difference soared to 20 p.p., as surveys alone suggest the Top 1% share has been declining for a quarter of a century.

Figures 3.18, 3.19 and 3.20 present analogous series for the Top 10%, Middle 40% and Bottom 50% income shares, respectively. The differences in both trends and levels between the national and fiscal income series vanishes progressively as we go from the top to the bottom of the distribution. For the Bottom 50%, both series are almost indistinguishable.

[56] Social contributions have increased steadily over time from 8-10% of national income in the 1970s and 1980s to 14-15% in the later 2000s, which largely fall on the income of middle classes. Similarly, undistributed profits of domestically-owned corporations have increased from 3-4% to 6-8% over the same years, but these are largely attributable to the highest earners. These evolutions are consistent with the notable expansion of the Brazilian welfare state after re-democratisation on the one hand, and the mass privatisation of state-owned companies and the growing financialisation of the economy on the other.

Figure 3.17: Top 1% Income Shares by Income Concept in Brazil: 1976–2016



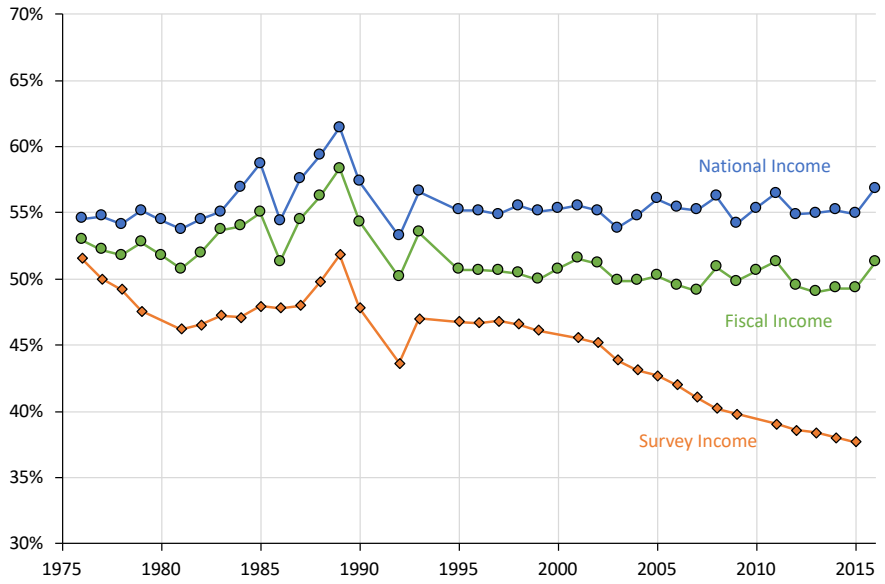
Notes: Distribution of pre-tax income among equal-split adults across income concepts. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data). Survey income shares use raw survey data. Fiscal income shares combine surveys and tax data. National income shares combine fiscal income from surveys and tax data with non-fiscal income from national accounts.

Likewise, surveys are increasingly more reliable as we go down the distribution of income, but this is almost by construction, since they provide the raw data for the poorest deciles for all three series. Even so, some important divergences remain. In levels, surveys also underestimate the income share of the Top 10% and overestimate the share of the Middle 40% and the Bottom 50%. This is corrected using the method of Chapter 1. Trends do not add up either: the survey figures show the Top 10% losing ground with both the Middle 40% and the Bottom 50% make substantial inroads, while the other two series show stability for the Top 10%, decline for the Middle 40% and smaller gains for the Bottom 50%.

Finally, Figure 3.21 compares Gini coefficients estimated from the three income concepts. Again, the fiscal and national income series yield similar results. Although the former suggests a lower Gini than the latter, the difference is typically just 3%-5% and trends do not deviate much, as both show the Gini falling by 12% to 14% since 1988 and by 4% to 5% since 2001. Yet, the survey series paints a very different picture: inequality levels are much lower and the fall in inequality is much larger – about 22% since 1988 and 13% since 2001. In general the Gini seems to mirror the inequality dynamics related to the Bottom 90% of the distribution, with middle and bottom shares falling during crises, such as those

of the early 1980s, early-mid-1990s and the recent recession from late-2014.

Figure 3.18: Top 10% Income Shares by Income Concept in Brazil: 1976–2016



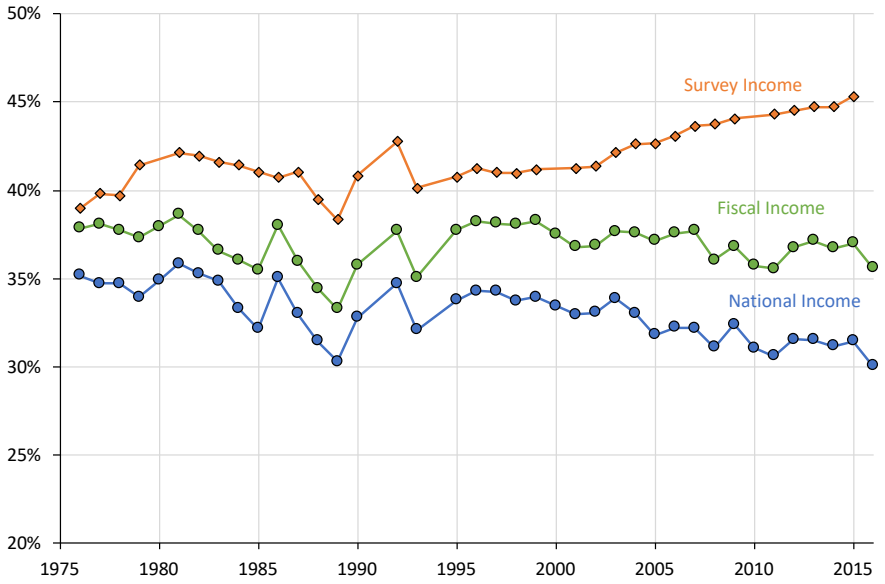
Notes: See Figure 3.17.

These results are consequential not only for the empirical analysis of income inequality in Brazil but also for more general methodological debates. In Brazil, worsening income coverage is misleading surveys into overstating the degree of inequality reduction. As useful as they are, surveys are rendering the rich increasingly invisible.

The main issue seems to be that cumulated income growth was not well captured in surveys since the 1970s, which ultimately gave way to the growing divergence between survey estimates of inequality and estimates combining multiple sources. Even if one declines to fully believe in Brazilian national accounting, and the growth it measures, the results from combining just survey and tax information strengthen our conclusions. Indeed the greatest part of the correction to the surveys, in terms of levels and trends, comes from the incorporation of fiscal incomes reported by tax returns to the top of the survey distribution of fiscal income. The resulting totals are not only closer in levels to those reported in national accounts, but growth rates become more similar (see Figure 3.4).

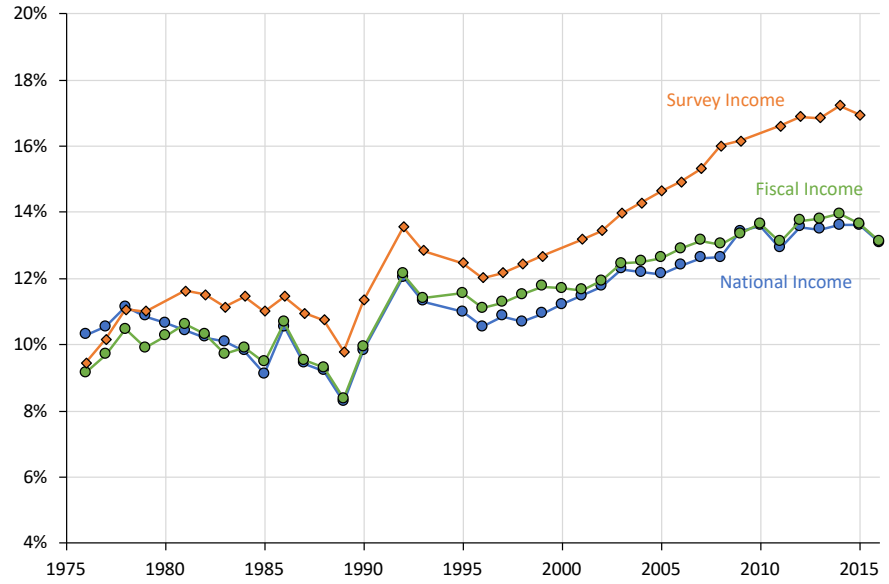
This growing fragility of surveys is a serious cause for concern. Researchers have known for decades that high incomes, especially from financial investments and business activities, are generally underreported in surveys (Hoffmann, 1988; Weinberg et al., 1999; Kennickell, 2009; Canberra Group, 2011). Such poor income coverage does not necessarily downward bias estimates of relative income differences, as these may theoretically be similar to figures derived from more complete data (Deaton, 2005). Nonetheless, there is mounting evidence that income coverage is falling over time in tandem with survey response rates, and that

Figure 3.19: Middle 40% Income Shares by Income Concept in Brazil: 1976–2016



Notes: See Figure 3.17.

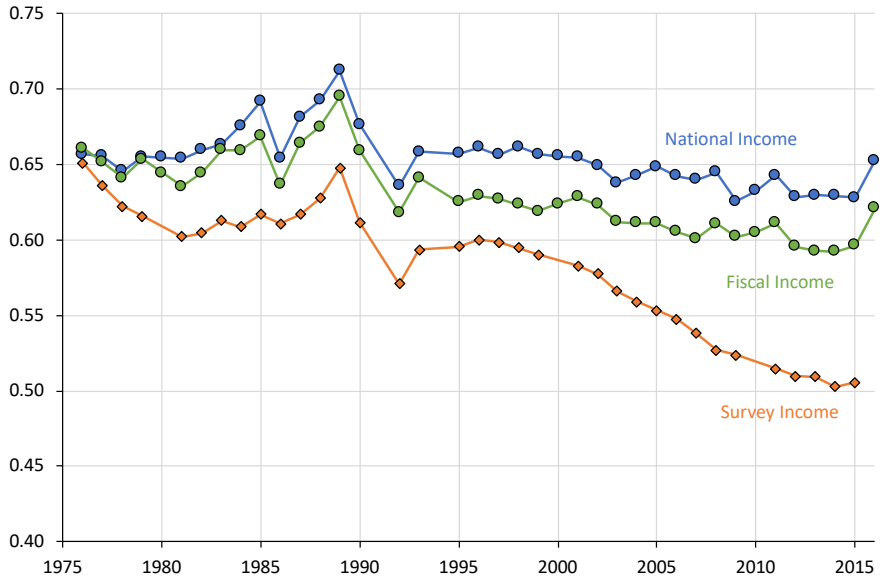
Figure 3.20: Bottom 50% Income Shares by Income Concept in Brazil: 1976–2016



Notes: See Figure 3.17.

this drop does not affect all income sources equally, thus possibly distorting both scholarly and political assessments of inequality (See Figure B.10 in Appendix B). Brazil offers a textbook example of this bias. According to surveys, persistently high inequality has been

Figure 3.21: Gini Coefficients in Brazil: 1976–2016



Notes: See Figure 3.17.

on the decline for two or three decades now, and the Top 10% income share plummeted from its peak of 52% in 1989 to just 38% in 2015. In other words, re-democratisation entailed a massive redistribution of income that could make the country unrecognisable in another 20 years or so if these patterns were to continue.[57] A careful assessment of the results from a more holistic treatment of the data proves this optimism unwarranted: although the Gini coefficient fell modestly and the Bottom 50% slightly increased their income share, this happened at the expense of the Middle 40%, not the Top 10%, which still gets about 55% of national income. In addition, the share of the Top 1% has been strongly rising in recent decades.[58] While focusing on total income may miss some interesting dynamics at play, at least it is less susceptible to shifts in remuneration behaviours linked to legislative and other temporal changes.

Finally, our results also stress the manner in which distinct analytical lenses can lead to conflicting conclusions about the same phenomena. This should call for more encompassing, multifaceted approaches. Scalar inequality indices summarising the entire distribution of income can be as concealing as they are revealing because they are much less capable of imparting the magnitude of the relative inequality in living standards than income shares. When the latter are based on raw surveys, or even raw tax data, they fail to take into account all registered income dynamics occurring over a calender year.

[57] See Alston et al. (2013) and Soares (2008) for buoyant interpretations along these lines.

[58] This is not to neglect that absolute material standards have improved with the fall of poverty levels registered by the surveys. Nor is it to overlook equalising tendencies made to the distribution of certain types of income, such as labour income and salaries. See chapter 2 for more details on this evolution.

5.2 Historical Analysis: Structural Change vs. Institutions

In light of the above exposition, how are we to understand the long-run evolution of inequality in an emerging economy such as Brazil's? And how does it relate to theories of growth and inequality, as presented in section 2? In this section we assess the institutional and structural factors behind the evolution of inequality and growth in Brazil.

To briefly summarise, "fundamental" economic and social structures have constrained inequality at a high relative level in Brazil, but it is the country's institutional history that is strongly correlated with its ebbs and flows. Brazil's development trajectory is derivative of the problems of late-development – the swift drive to modernisation nesting between competing visions of different social actors about how the proceeds of production should be distributed for this end; what the government's role in this drive should be; and how to reconcile international trends in economic policy with domestic priorities. These multiple dilemmas inevitably gave way to distributional bottlenecks, that were often overcome by force rather than by legal-democratic means.

5.2.1 Pre-Modern Foundations

The emergence of high inequality in Brazil has its origin in a given set of historical contingencies, both foreign and domestic, from which particular socio-economic structures and accommodating institutions arose. These relate to the early geographical aspects of wealth (land, migrant settlement) and an elitist form of governance and decision-making, inherited from the colonial era (c. 1500–1822), leading over time to the regional cleavages in economic capacity and political power. This was exacerbated by the distribution of land, slavery and the export trade, considered by scholars to be the key institutions of Brazilian society for over 300 years (Dean, 1971; Engerman and Sokoloff, 2012). Land concentration and agrarian reform have been perennial issues in the history of Brazilian development, most notably because the structure of rural property was so tied to slavery and the export trade.

Brazil was a predominately rural country until the 1970s – the share of the rural population gradually falling from 90% in 1890 to 44% by 1970. The best estimates of land concentration for Brazil date from 1975, when the first agricultural census was conducted. By this year the Gini of land ownership, estimated by Hoffmann and Ney (2010, Table 4) for the entire country, was 0.85 (substantially higher than our national income Gini of 0.65) – a level it has since not fluctuated from. The highest levels of land inequality are observed in the states of the northeast. Extreme concentration can be observed from the fact that the top 5% largest landholdings have accounted for 69% of total holdings since 1975, while the bottom 50% smallest landholdings have represented just 2% (Hoffmann and Ney, 2010, Table 5). The persistence in the levels of concentration illustrates the resilience of this institution. We can only speculate that land ownership was at least as unequal – probably more unequal – in the preceding 100 years. Indeed, rural poverty and inequality were the

driving factors behind Brazil's rural exodus and rapid urbanisation rates across the 20th century, as agrarian workers flocked to cities in search for higher paid and better protected urban jobs.

Similar to other countries in the region, the combination of these endowments, institutions and disruptive contingencies (disease and mortality crises, de-colonisation, international commodity booms) formed the structural problems of late-development – "late" in large part due to this cocktail of factors (Williamson, 2011, 2015). However, this trajectory does not imply, as Williamson (2015) notes, that initial endowments and foreign colonisation alone made Latin American countries, like Brazil, more unequal than other regions around the world. In fact, he shows that "income inequality in pre-industrial Latin America (pre-1880) [was] lower than pre-industrial northwest Europe (pre-1800) and the early industrialising United States (1860)", based on Ginis calculated from occupational censuses and tax data, and ratios of average income to unskilled wages. What seems to have made Latin American nations more unequal in *relative* terms was the continued *belle époque* rise over the 1910s–1970s period. Our top income share estimates confirm this, as the share of the Top 1% in the mid-1920s was similar to that of developed countries, like France and the USA (see Figure 3.13), only becoming relatively higher after 1929.

The drive to hastily expand productive capacities and material standards of living in Brazil during the 20th century brought structural change into conflict with institutional transformations deemed necessary for the continuous development of the economy, such as the role of the state in resource allocation and pricing policy, as well as in the mediation of class relations. As the country became richer, economic surpluses could increase, and so could the extractive capacity of elites. The degree to which old (land-based) and new (industry-based) elites could maintain their extraction rates depended on their political influence in co-opting these future institutional changes. Their relative proximity to the developmental objectives (and associated theories) of policymakers proved to be lucrative for economic elites over the long-run. But how was the mediation between institutional transformation and structural change played out in Brazil, and precisely why did the country miss out on a more sustained "20th Century Egalitarian Levelling"?

5.2.2 Why did Brazil not Sustain its 20th Century Levelling?

To begin with, we can relate our estimated inequality trends with other long-run indicators. Figure 3.22 shows capital and labour shares alongside our Top 1% share in national income. In general, the elasticity of personal income inequality (as measured by income concentration) to the capital share seems to be quite high.[59] This correlation is consistent with the findings of Bengtsson and Waldenström (2018) for other countries. In Brazil the correlation is most emphatic between between 1945 and 2005. The key contributing factor

[59]It is typically inferred that the returns to capital are more unequally distributed than those associated to labour.

to this dynamic is arguably the increasing wedge that emerged between productivity and real wages since 1964, compared to the years 1950-1964, when average wages were equal to or slightly outpacing productivity (Figure 3.23), driven by the sharp re-valorisations the minimum wage over the period (Figure 3.24).

Outside of the 1945-1950 period, between-group inequality (between wage earners and capitalists) was probably less important than within-group inequality. In the inter-war period, the rise of skilled labour, especially after the Vargas-led "urban revolution" took power in 1930 (Furtado, 1965), paralleled with the steady decline of agriculture in value-added and increases in rural-urban migration, can explain the increased concentration of income through wages.[60] This is very much in line with Kuznet's (1955) hypothesis. During the 2000s, compositional effects likely dominated. Both labour and capital remuneration increased in real terms, but the former slightly more than the latter (as evidenced from Figure 3.23). This was especially due to the rise in median and bottom wages at faster rates than the average, anchored to strong minimum wage growth (see Figure 3.25).[61] But concentration grew as average top incomes outstripped the average total income for the full population, in the context of a booming economy (as shown in Figure 3.24).[62]

Interestingly, these dynamics do not fully conform to the strong linear rise in inequality posited by Williamson (2015) for Latin American countries (including Brazil) during the 1920s-1970s period. Rather, in the Brazilian case, a wave-like pattern is observed, not dissimilar to the one experienced by Argentina (see Figure 3.14). This trend reveals the important role that institutional changes, which the "planning" of structural change was fomenting, had in shaping the Brazilian income distribution. As briefly noted in section 4.1, important political transformations swept the institutional landscape: from Vargas' urban-backed revolt in the 1930s and "corporatist" *Estado Novo* dictatorship (1937–1945), to the growing politicisation of labour unions, and their influence on government wage policy; to the renewed electoral-democratic pressures in the subsequent two decades under the urban-worker alliance of the Brazilian Social Democratic Party (PSD) and Labour Party (PTB). These changes were very much endogenous to the economic cycle, which was initially driven by external factors (international demand fluctuations, World War II), and then determined by explicit state management.

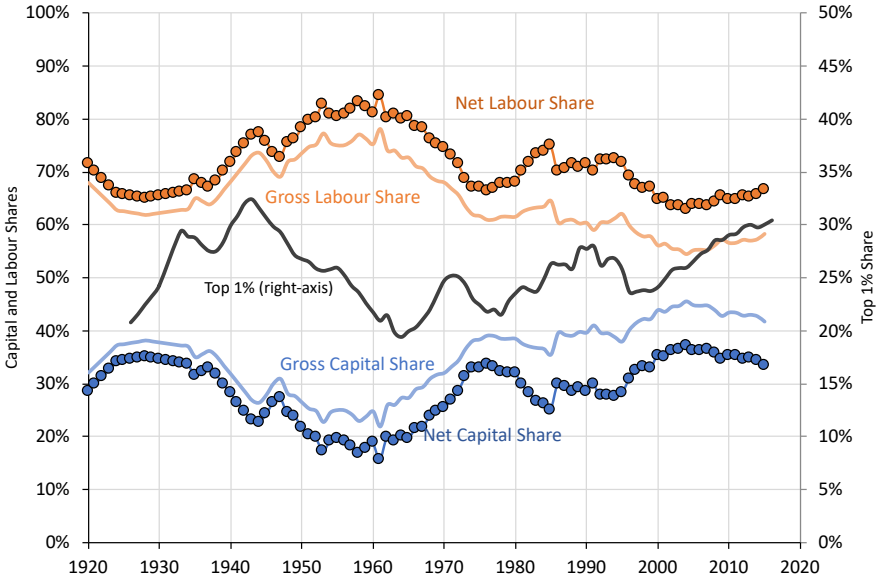
The initial actions of the federal government to industrialise the country were consistent with the view of stimulating private capital through pricing/output protection and long-

[60] As can be seen from Figure 3.24, the real minimum wage receded notably between 1940 and 1950, while average total income and Top 1% average income advanced.

[61] Between 2005 and 2013, the year prior to the domestic crisis, average total income grew by 18%, median income grew by 28% while the federal minimum wage rose by 31%. Extending the time-frame to 2016 reproduces the same ranking with more magnified differences – the real value of the minimum wage was now 35% higher, median income was only 9% higher, while average income was 2% higher. For more details on this period see chapter 2.

[62] The mixed income share fell from 12% in 2000 to 6% in 2015. Excluding mixed income, the net capital share remains pretty stable over the period (falling by -0.6 points instead of -2 points), while the labour share increases by a faster rate (by +6 points, instead of +2 points).

Figure 3.22: Factor Shares vs. Top 1% Share in National Income: 1920–2016

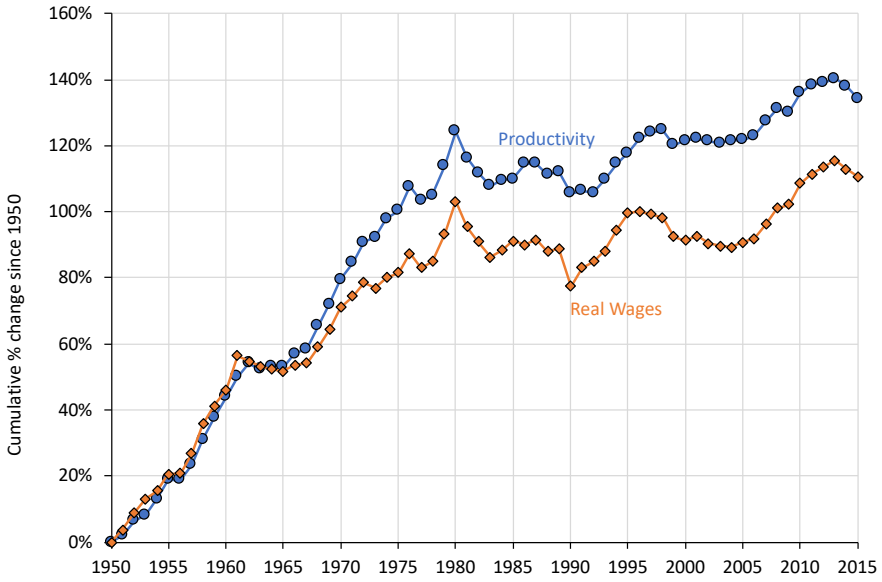


Notes: Factors shares in national income. Net shares are after deduction of capital depreciation. Authors' estimates for 2000-2015 using data from IBGE (2017). We divide mixed income between labour and capital according to a 70–30 split. Pre-2000 estimates are anchored to the annual change in the labour share computed by Frankema (2010) for Brazil. Top 1% shares are 3-year moving averages of the series in Figure 3.5.

term credit. Obstacles of scale and complexity (as well as security concerns during the war) would however push the government into direct production and control, as part of specific objectives to upgrade productive capacity. Thus, over the 1930s and early-1940s the government's price and output management of productive sectors (mainly coffee, but also other commodities) would eventually be followed by direct public production from scratch (steel) or via outright nationalisations (shipping) (Baer, 2014). While state management was becoming ever more central, the importance of WWII for the Estado Novo increase in inequality should not be overlooked (see Figure 3.5). The war allowed Vargas to push for worker exploitation in the so-called "war industries", with the excuse of "helping the nation", while at the same time maximising profits for industrialists. Since businesses could not import capital goods to increase production, the existing capacity had to be taken to its limit. The government complied by imposing rigid discipline. Wages were kept low, while longer and more intense shifts were mandated in several industries (Paoli, 1989; Gomes, 2005). Furthermore, disruptions to international trade sheltered domestic producers from foreign competitors and even opened up opportunities for Brazilian exports, especially in textiles and other manufacturing goods, causing a short-lived boom (Abreu, 1990; Baer, 2014).[63] Thus, contrary to developed countries, the years of the war were not

[63]Imports fell from almost 12% of GDP in 1938 to 7% in 1945, and then rebounded in 1945-

Figure 3.23: Evolution of Productivity vs. Real Wages since 1950



Notes: The graph shows the cumulative percentage change in productivity and real wages since 1950. Author's computations using various sources. Productivity is defined as GDP per employed worker using data from IBGE (2006, 2017). Real wages correspond to total labour compensation per worker, using labour income from Figure 3.22 and employment data from The Conference Board Total Economy Database. All incomes are deflated by the GDP deflator from IBGE.

met with a fall in income concentration (see Figure 3.13).

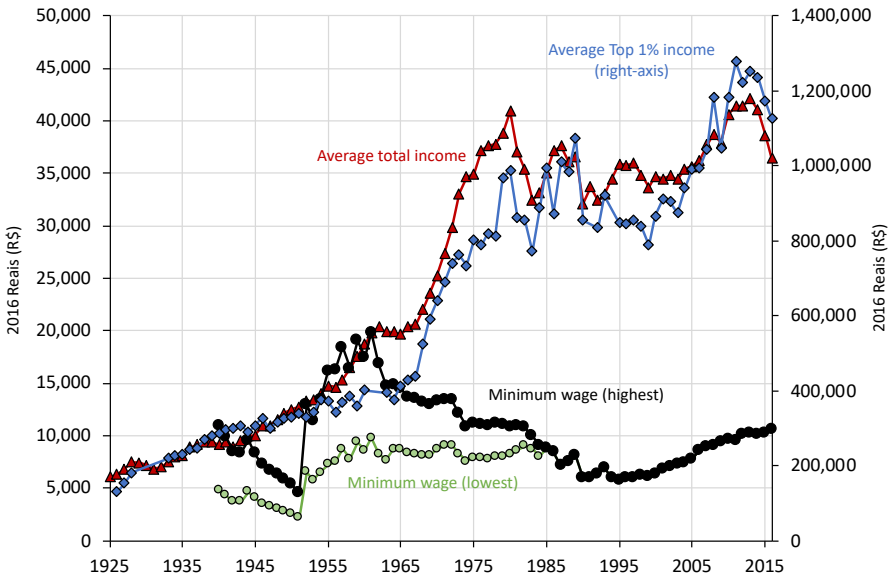
During the 1950s, state ownership of firms expanded rapidly, notably in machinery, mineral refinement, railroads, public utilities, transport, communication and banking. Concerning the latter, the creation in 1952 of the country's first national development bank (the BNDE) marked a significant turn in government policy. By financing large infrastructure projects, as well as capital injections into heavy industry in exchange for equity participation, the government became a "reluctant owner of enterprises" (Baer, 2014).[64] While the government increasingly embraced a "developmental-nationalist" strategy, it was by no means aiming for a Soviet-style takeover of the economy (Skidmore, 1976). Rather, it had specific "missions" on the agenda that included private and foreign participation, with the government acting as producer of last resort. This was typified in the *Programa de Metas* of the Kubitschek administration (1956–1961).[65] Our results show that during this

1947. Exports were far more stable at 11%-12% of GDP, with a slight downward trend after 1945. Manufacturing exports, however, follow an abrupt inverted-U shaped trend, from almost non-existent in 1938 to 2% of GDP in 1943 and back to irrelevance after 1945 (see IBGE, *Anuário Econômico do Brasil 1952*, p. 285).

[64]Government involvement was necessary given the limited resources of private sector to finance projects deemed cornerstones of Brazil's industrialisation plans. This was the case, for example, of four of the country's most important steel firms, which passed into government majority control via the capital injections of the BNDE and the Banco do Brasil (Baer, 2014).

[65]The "Missions Program" was a plan elabourated by the new Development Council of the

Figure 3.24: Evolution of Average Incomes vs Minimum Wages: 1940–2016



Notes: Average Top 1% income relates to distribution of pre-tax national among equal-split adults across income concepts. Authors' calculations (combining survey, tax and national accounts data). Average total income corresponds to national income per adult. Minimum wage data is from the Ministry of Labour and Employment (MTE). The first minimum wages were introduced in urban areas at the state-district-level. A minimum wage for rural workers was introduced in 1963. The national minimum wage was introduced in 1984. Before 1984 the series reports the highest state-level minimum wage observed in the country (the Federal District of Rio de Janeiro (city) until 1961, then Rio city and the new Federal District, Brasilia, until 1964, then the two alongside São Paulo city until 1984). The lowest minimum wage corresponds to the district average of the state of Maranhão in the Northeast, also the state with the highest land Gini in the country (Hoffmann and Ney, 2010, Table 7). All incomes are deflated by the GDP deflator from IBGE.

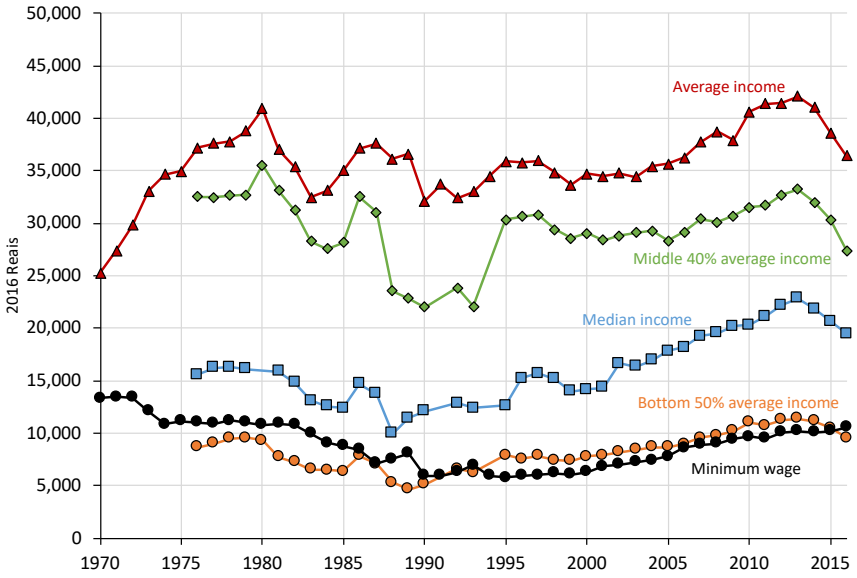
whole period growth was distributed more favourably to the average worker outside of the Top 1% (see section 4.2). Sharply rising wages, at or above productivity levels, anchored by hikes in urban minimum wages (granted to expand internal demand and protect unionised workers from rising costs of living) were driving the trends.[66]

From the 1930 "Revolution" to the early 1960s public views on inequality were mainly expressed in terms of the "social question" (*"questão social"*), framed as rights and responsibilities of labour versus capital in the growth process. This question was largely neglected during the Old Republic. Its liberal ideology was deemed bankrupt in Vargas' "corporatist"

President in 1958, which specified investment and production objectives in the areas of energy, transport and basic industries, as well as (on a secondary plain) in agriculture and education. It formed a concrete part of the president's motto of "50 years of development in 5" (Skidmore, 1976).

[66] It should be noted that the minimum wage only covered formal urban workers, excluding rural formal workers (until 1963), and informal self-employed workers and domestic servants. Thus, the official minimum wage was not quite a wage floor, but rather an indicator closer to median income, at least until the mid-1970s (see Taylor et al. (1980) and Figure 3.25).

Figure 3.25: Evolution of Real Income by Group in Brazil: 1970–2016



Notes: Distribution of pre-tax income among equal-split adults across income concepts. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data). Median income corresponds to the income at the 50$\hat{\text{t}}$h percentile. For minimum wage data, see Figure 3.24.

turn. This was conceived of as a third-way between market liberalism and centrally-planned communism. From 1943, anticipation of the end of the war and its favourable conditions meant that Vargas was no longer motivated to co-opt and control the masses, but openly sought to spur mobilisation for new sources of legitimacy. The government thus elevated the status of the (formal) "worker" and enacted new progressive work-related laws, seeking a more productive collaboration between capital and labour (including revising wage policy and passing the consolidation of Labour Laws (CLT) in 1943 – the latter are to this day the main statute regulating labour rights in Brazil). The social question only gained more traction, as income concentration commenced its downward secular cycle. In his Labour Day speech of 1944, Vargas proclaimed that:

> "liberty, in the strict sense of political franchising, is not enough to solve the complex social question...Supporting workers economically is tantamount to giving them the true sense of freedom and security to express their political opinions. And for this, it is urgent to correct the imbalance between those who find no limits in the profitable exploitation of the means of production and those who toil in a permanent state of necessity, without resources to acquire what is indispensable for survival" (Vargas, 1951, p. 291).

In 1946, a year after his removal from power by the military (fearing continued leftist "populism"), Vargas would pronounce the passing of the "old liberal and capitalist democracy",

which was "founded on inequality", and the coming of "a socialist democracy, a workers' democracy" (Skidmore, 1976, p. 107).

The return of electoral democracy, along with heightened voter participation, ensured that this line of thinking would be broadly continued in the 1950s, first during Vargas' term as elected president (1951-1954), and after by members of his two political creations: the Social Democratic Party (PSD) and the Brazilian Labour Party (PTB). In 1952 the minimum wage adjustments granted by Vargas recouped all the previous losses due to inflation since 1940, as the real value of the minimum wage increased by 160% on average across all states from 1951 to 1952 (see Figure 3.24 and Figure B.18 of Appendix B). But now more than ever, social problems were becoming more and more associated to the country's low industrial capacity. As before, raising living standards (of those only depending on their labour) was seemingly more primordial than explicitly addressing inequality *per se*, except that greater impetus was given to development missions to achieve this. This was especially the case under the Kubitschek (PSD) presidency. The social question was not about whether wealth was in public or private hands, but what private capital owners did with their surpluses (which the government would help to foment). As long as the rich productively invested the best part of their surpluses, they were doing their patriotic duty. In his address to the national congress after his election in 1956, Kubitschek proclaimed that:[67]

> "consumption austerity of the most favoured classes is not only an essential condition to accelerate capital accumulation and an example for the classes that ascend in the social ladder; it is also a political objective capable of strengthening the doctrine of freedom and private initiative. The disparity of consumption patterns within a society can be a stimulus to social mobility, if not exaggerated, but it will certainly be a cause of class struggles when taken to extremes, common in less developed countries."

Not only was it a duty of capitalists to invest, but given time constraints and perceptions of social injustice in the "cumulative concentration of resources", "[s]tate intervention aimed at accelerating reproductive investments and creating an austere consumption discipline, becomes a logical imposition in regions that, like Brazil, are now in the initial stage of the process of economic development" (ibid). Again, like Vargas, the aim was not complete government takeover of the all economic resources, but rather to evade the "communist threat" by intensively stimulating national production and wealth to the benefit of workers.

As seen from Figures 3.1 and 3.2, the 1950s were an exceptional period in Brazilian developmental history. Per adult income grew by 4.1% per year over the decade, driven by the most intense period of industrialisation the country has ever experienced (manufacturing increased its share in value added by 5.2% per year in the seven years between 1952 and

[67] Juscelino Kubitschek, Mensagem ao Congresso Nacional, 1956, p. 276 (authors' translation). http://www.biblioteca.presidencia.gov.br/presidencia/ex-presidentes/jk/mensagens-presidenciais/mensagem-ao-congresso-nacional-jk-1956-parte-1.pdf.

1959).[68] Formal employment registers made by the Ministry of Labour and Social Insurance increased by almost 70% between the years 1948-1950 and 1958-1960, as the urban share of population increased from 36% in 1950 to 45% in 1960 (IBGE, 2017). With this, regional inequality was increasingly becoming an important issue among government circles. This was typified in Kubitschek's election pledge to construct a new capital city from scratch in less than 5 years in the centre of the country, away from the politically and economically dominant southeast.[69]

In a coordinated effort to promote "progressive industrialisation", the government's planning of structural change fomented institutional changes that would be the cause of its own downfall. With the continuous growth of the urban proletariat, associated unionisation grew by 5.6% increase per year between 1938 and 1963 (IBGE, 2006), with urban union density increasing from approximately 7% to 15% among all urban workers and from about 11% to 24% among urban employees.[70] Average wages grew fastest in the more urban-dense regions of the south and southeast, under an executive that was increasingly dependent on the urban vote.[71] These regions also concentrated regional union density, with the southeast alone comprising 75-80% of all union members during this period (IBGE, 2006).

The growing urban–rural divide was a direct implication of the electoral system and of regional class structures, which largely hinged on the literacy of the population. While the political executive (the president and state governors) was elected by majority voting by literate citizens – who were more concentrated in urban areas – members of the legislature (i.e the congress) were elected by proportional representation (Furtado, 1965). With the majority of the population still residing in rural areas during this period, the literacy condition "made a given vote in a rural congressional district (sometimes with as little as 10 per cent literacy) much more valuable than in the cities" (Love, 1970, p. 22).[72]

Thus, the un-addressed cleavage between the urban-orientated "populist" executive and the rural-dominated "patriarchal" legislature was increasing its stranglehold over the planning of development. This was an important source of the ensuing political instability of the early 1960s. It was compounded by the fact that the PTB – the political-wing of

[68] This was greater than the increase during the Brazilian "economic miracle" years between 1967 and 1974 (4.1% per year). See Figure 3.2.

[69] Estimates by Morgan (2015) for the 1945-1950 period reveal that income concentration was about 3 times higher in the more rural northern and western states than in more urbanised southern states (Figure C.3).

[70] Authors' computations using union data from IBGE (2006) and employment data from IBGE (2017). All urban workers include employees, employers and liberal professionals/self-employed in urban economic activities. We define total "Urban employees" as the active workers in industry, commerce, transport and communication, and public administration. This denominator also includes the employers in these activities, as well as any self-employed, which we cannot isolate. If we just restrict the total urban employee denominator to workers in industry and public administration the union density rises from about 18% in 1938 to 40% in 1963.

[71] See Figure 3.24 and section A.9.3 in the Appendix for the discrepancy in regional wages.

[72] Representation in the senate (the lower house of the legislature) also functioned along similar lines. In addition, the Constitution allowed congress to impose limits on the decree powers of the president, which only further escalated tensions between the two demographics (Furtado, 1965).

the urban unions – took over the presidency in 1961, and made the most gains out of any party in the congressional elections the following year.[73] This was mostly a result of the strategy of mass mobilisation of urban workers initiated by Vargas after 1943 and propelled in the 1950s by the PSD executives and the PTB-controlled Ministry of Labour: between 1946 and 1961, the share of voters practically doubled from 20% of the voting-age population to 40% (IBGE, 2017). Under President João Goulart the "social question" was increasingly subsumed under the need for "*reformas de base*" ("fundamental reforms") , that is, wide-ranging reforms that could both spur growth and address social needs at the same time. The central pillar of these changes was the thorny issue land reform.

In march 1963, a radical proposal was presented in congress to change the *latifundist*-based rural property regime to overcome structural bottlenecks to further industrial expansion. This involved government expropriation and redistribution of lands, with compensation taking the form of government bonds instead of up-front cash, as stated in 1946 Constitution. This was founded upon the idea that the archaic system of rural property prevented the necessary expansion of agricultural production, which in turn would constrain financial expansion and thus the market for manufactured goods in the countryside. This proposal met with huge opposition in congress, where powerful land-owners wielded large influence from the over-representation of rural districts. (Skidmore, 1976).[74] In the same year, a rural labour code was finally implemented (30 years after its urban counterpart), granting rural workers the right to a minimum wage defined by state, like its urban-equivalent, and creating a rural workers confederation (CONTAG). Another significant reform related to foreign capital. A law restricting the remittance of the profits of foreign companies (approved by congress in 1962) was emitted by Goulart in January 1964.[75]

To help push through these reforms, Goulart's executive increasingly bet on mass mobilisation. By 1964, after a successful referendum result in the previous year that gave Goulart executive powers as president – repealing the 1961 Constitutional Amendment that abolished the executive role of the presidency and created a parliamentary system – the president broke with the historical compromise between capitalists and workers that defined the post-Vargas era, to much political tension (Love, 1970; Skidmore, 1976;

[73] In the 1945 legislative elections the PTB obtained 10 per cent of seats in congress. By 1962, this had shot up to almost 50% (Love, 1970).

[74] In May of 1963 a congressional commission rejected the proposed land reform by a majority of 7 to 4 (Skidmore, 1976).

[75] This law defined the foreign capital from which profits could be remitted as excluding reinvested earnings in the Brazilian territory. These would be treated as "national capital", to the strong opposition of foreign investors, and particularly the United States (which accounted for around 40% of total FDI through the 1950s). (Skidmore, 1976; Baer, 2014). US ambassador to Brazil, Lincoln Gordon, heavily deplored Goulart's direction and warned of a Peronist-style takeover of the political scene, in a meeting of the Foreign Relations Committee of the US Senate (Skidmore, 1976), while the *New York Times* strongly attacked the law in an editorial in January 1964 (see `https://www.nytimes.com/1964/01/27/archives/brazilian-boomerang.html`.

Weyland, 1996).[76] At a massive public demonstration ("Comício da Central") in Rio de Janeiro, in March 1964, a couple of weeks prior to the military *coup*, Goulart announced he would implement two main decrees (now that voters had granted him executive powers): the nationalisation of private oil refineries and the expropriation of all privately-owned lands above a certain size, which were located in areas of strategic national interest.[77] The president declared further planned decrees, including rent controls; and congressional bills, including universal suffrage, i.e. giving the illiterates the right to vote (ibid). (ibid).

The post-Estado Novo compromise worked well when the economy was growing at over 10% per year on average since 1945, but faced heavy strain when the economy began to slow down in the early 1960s. This proved to be the final straw for the "workers' democracy". This was exacerbated further by the military's real fears that an impending "sindicalist" workers' movement could replace the armed forces as the most powerful group in the Brazilian political scene (Skidmore, 1976). The fear was fuelled by the creation, in August of 1962, of the "General Command of Workers" (CGT) by the worker confederations of all economic sectors (including the rural-workers confederation after 1963, but excluding commercial sector workers), as a management organ to promote further horizontal integration of the syndicalist movement. It was immediately extinguished after the military takeover in April of 1964 (Costa, 1981).

The "crisis" of the early 1960s brought about a reactionary civil-military *coup* to re-orientate the country's priorities and development strategy.[78] The rural-urban cleavage would take a new turn during the military regime, as it abolished the urban-biased elections for the president of the republic and state governors for the large part of its reign. Thus the regime only permitted electoral "competition" for elections of federal and state congressional deputies (which as explained, depended more on traditional rural society). In these regional and local elections, the military party (ARENA) "competed" with the artificially-created Brazilian Democratic Movement (MDB), after having banned all previously existing parties. Given this political system, the regime's sustainability depended largely on overall industrial-growth and relative living standards in the more underdeveloped regions. Minimum wages were more stable/fell less sharply in states of the northeast, but also in the north and centerwest regions (see section 9.3 of Appendix B). This significantly increased the average rural-urban wage ratio, which went from 58% in 1966 to 71% in 1973 for workers outside the state of São Paulo, and from 48% to 117% for workers in São Paulo state (Taylor et al., 1980, Table 10-8 and Table 10-9).

[76] This is manifest from his increasingly radical and rhetorical speeches. See `http://funag.gov.br/loja/download/641-Discursos_joao_goulart.pdf`.

[77] Specifically lands exceeding 100 hectares, when they were located within 10 kilometres from federal roads and railways; and lands exceeding 30 hectares, when situated in areas constituting irrigation basins of federal dams (Skidmore, 1976).

[78] This intervention was closely monitored by the United States. See Appendix in Skidmore (1976) and `https://nsarchive2.gwu.edu/NSAEBB/NSAEBB465/`.

5.2.3 Competing Interpretations of the 1960s Crisis and Inequality Trends

The key role of institutions in determining the rates of growth of relative incomes is nowhere more evident that during the years of the military dictatorship (see section 4.2 and Figure 3.24). The legacy it created can be seen in the decades following its collapse, as Figure 3.23 on the productivity-wage gap clearly depicts. Therefore, it is paramount to assess more closely the ideas behind the paradigm shift of the 1960s.

Clashing interpretations of the crisis of 1962–1965 produced new conventions regarding government's distributive-growth policy, and new inequality trends, which our results illustrate. On the one side of the "crisis debate", there were those who argued for the structural "stagnation" thesis, most notably Ceslo Furtado, former Minister for Planning in Goulart's government (see for instance, Furtado (1969)). This thesis posited that inequality itself was a bottleneck for greater expansion and development. This was premised on the nature of Brazil's import-substitution industrialisation (ISI) since the 1940s and on domestic elites' seeking to emulate consumption patterns of their developed-country counterparts. ISI targeted capital-intensive durable goods, for which skilled workers would be required, while the country maintained a large pre-capitalist primary sector with unlimited labour supplies. The resulting inequality and excess capacity would squeeze capital productivity and lead to structural unemployment and a fall in incomes.

On the other side, the crisis was theorised as a crisis of "investment", in the traditional Keynesian sense of low "animal spirits" (profit expectations), by equally dissident economists of the military regime, such as José Serra and Maria Tavares (for a summary see Serra and Tavares (2000)). While it was related to the depletion of a cycle of the ISI model, it was not caused by the reduction in capital productivity, but by the lack of investments that would complement the existing industry. New projects needed to be introduced in an adequate sequential timing, that is, after the maturity of the investments in the "Plano de Metas" of the Kubitschek government (1956-1961). The low business expectations were due to the political economy of the early 1960s – the inflationary wage policy of the alleged "syndicalist-Perónist" executive, and the agitation surrounding its reform proposals.

While both sides of the debate agreed that incomes were still too highly concentrated to best ration consumption-demand, Serra and Tavares argued that the fall in the surplus–wage relation over time constrained investment from the supply-side. Thus, the solution to the crisis lied in one of two paths: increasing the profit–wage relation by repressing labour costs, so that both falling investment and rising inflation could be tackled; or introducing new rounds of public mission-investments, while taking purchasing-power out of the top of the distribution through a reformed income tax or land redistribution, to control inflation. Unsurprisingly, the former was the dictatorship's chosen path. This was embodied in the newly-drafted Government Economic Action Program (PAEG) in 1964 and in the new 1967 Constitution. Everything was channelled to increase the surplus–wage ratio and quell inflation, from the general wage squeeze and capital-friendly tax reforms, to

cuts in government expenditures (Taylor et al., 1980; Serra and Tavares, 2000).[79] The Brazilian economic "miracle" (1967-1974) buried the stagnation hypothesis, in line with what Serra and Tavares expected from the ensuing economic and political climate. The official government line, repeated in the principal news outlets, was that living standards were improving across the board and that inequality would fall in due course – but this could not be rushed, otherwise inflation would jeopardise real gains (Souza, 2016).

A notable moment in the growth-inequality debate came upon the release of the 1970 Census tabulations on income. Brazilian economists, such as Hoffmann and Duarte, as well as foreign economists, most notably Albert Fishlow, quickly wrote papers showing that inequality had risen considerably in the 1960s, and linking the results to deliberate government policy, particularly concerning wages and spending (Hoffmann and Duarte, 1972; Fishlow, 1972). This became an acute political problem after Robert McNamara, then president of the World Bank, used Fishlow's (as yet unpublished) results to criticise the high levels of inequality in Brazil. Researchers and intellectuals aligned with the dictatorship scrambled for answers. The distribution of income thus became the focal point of charges against the "Brazilian expansionary model" (Souza, 2016).

The Ministry for Finance then commissioned academic Carlos Langoni to rebuke the claims made by Fishlow and others, granting him exclusive access to new data. Working off large micro-samples of both 1960 and 1970 census (as opposed to summary tabulations), as well as income tax tabulations, Langoni interpreted the rise in inequality as a combination of Kuznets-like structural change and the short-run in-elasticity of the supply of skilled workers (Langoni, 1973). This subsequently became the official line of the regime. In the forward to Langoni's book, fellow academic and Finance Minister of the 1967–1974 administration, Antonio Delfim Netto, gave institutional backing to Langoni's thesis that the rise in inequality resulted from the "market disequilibria accompanying the development process". He justified the chosen path of economic development by claiming that modern sectors pay relatively more because of their higher productivity, even if the dispersion of earnings is greater. Therefore, "it makes no sense to use the increase in inequality as an index of welfare deterioration" (Langoni, 1973, p.14). The skills-differentials hypothesis was also backed by the Delfim's successor, Mario Simonsen, who stated in a congressional study on the government's wage policy that "the main reason for worsening income distribution seems to have been an inadequate profile of labour qualifications vis-à-vis the needs of the

[79] After the 1964 *coup*, many labour leaders were arrested and some unions revoked. The new 1967 Constitution put severe legal constraints on strikes and all collective bargaining between labour unions and employer associations were suspended. Subsequently, all wage agreements were to be regulated by a wage formula dictated by the government, which was based on (i) reestablishing the average real wage observed in the past 24 months, (ii) maintaining during the following 12 months the same average real wage as in the preceding year, (iii) adjustments for productivity. In practice, expected future inflation was consistently underestimated until late 1960s, while the productivity factor was much smaller than the observed growth of GDP per capita through the mid-1970s. These were the main culprits of the effective wage squeeze (Bacha, 1980).

market", with skilled labour being rather scarce (Simonsen, 1975, p. 18).[80] This begs the question of whether remuneration practices are only determined by the skill-set of workers and changes in their demand (in line with neoclassical theory) or whether they are also rooted in changing institutional conventions (in the Keynesian/structuralist spirit).[81]

According to our results, a clear upward trend in the capital share and in income concentration followed the 1964 *coup*. This was especially the case over the 1960s, such that a rise in income concentration found by other authors at the time seems indisputable (see Figure 3.22).[82] However, the short-run decline in concentration we find during the early 1970s seems contrary to the studies pointing towards an increase in the wage spread over these years of the "miracle" (see Taylor et al. (1980) for a summary). Two hypotheses are possible. One is that the continuous rise in inequality, especially from labour earnings, was overstated. This seems in-and-of itself unlikely given that many of the conclusions regarding the wage spread were reached using fiscal data, particularly data from the so-called "Two-Thirds Law".[83] Incomes from this data source reveal rising inequality from 1969 to 1973 (Morley, 1978). The second hypothesis that could reconcile these results, is that the fall in inequality that we measure over early 1970s is due to high earners shifting income from wages to profits, given new fiscal incentives that were granted to capital income by the dictatorship. Across the late-1960s and early 1970s new deductions for capital incomes were created to incentivise private equity investments and federal government debt purchases. These included the receipt of dividends up to a certain annually variable amount. Moreover, the late-1960s tax reform gradually removed capital incomes from the progressive income tax schedule towards exclusively schedules with lower rates (Nóbrega, 2014).[84] Unfortunately, our estimates for top shares during this period cannot directly capture this shift (if it did occur), since they are anchored to gross taxable incomes, which exclude incentivised investment income (see Appendix B). But support for this hypothesis can be sourced from the fact that the capital share in the economy continued to grow

[80] Simonsen cites census data that reveal that in 1960 workers with a high school diploma comprised 10.3% of the labour force, while college graduates accounted for 1.6% (Simonsen, 1975, p. 18). As Morley (1978) argues, however, formal education is not a perfect synonym for "skills", but rather for "trainability". Thus, the skills relevant for industrial production come from "experience", not necessarily from college education. This implies that, in a booming economy, institutional practices regarding in-work remuneration and promotion may matter more than supply-side variables.

[81] See section 2.

[82] The difference is that the levels of inequality we estimate are substantially higher than the census-based estimates, given the the discrepancies between survey-based and fiscal-based sources. For instance Langoni's Top 1% share in total individual income moves from 12.1% in 1960 to 14.6% in 1970. His estimate from the tax tabulation of 1970 is 10.5%, which corresponds to the share of the Top 1% of tax filers not the Top 1% in the total population. Our share for the same income concept for the total population is 15% (see Appendix B, section 5.4).

[83] This law required firms to annually submit information to the Ministry of Labour to prove that at least two-thirds of the workers on their payroll are Brazilian citizens. Among the information provided were the names, citizenship and monthly wages of all the firms' employees.

[84] This is evident from the gap we measure between taxable income shares and total income shares (see section 5.4 of Appendix B).

beyond 1970, at least up to the late-1970s (see Figure 3.22).[85]

In any event, the distributional debate of the early 1970s had notable effects on the dictatorship's outlook on inequality. The first two stages of the dictatorship were more concerned with controlling inflation, initially by tight fiscal and monetary policies (1964–1967), and then through rigid wage policy and passive monetary policy (1967–1974). Contrary to the first National Development Plan (1967–1974), the regime's second Plan (1975–1979), in the midst of the inequality debate, explicitly mentioned the issue, claiming that "growth cannot resolve the problem of the adequate distribution of income, if left to the simple evolution of market forces. And on the other hand, the solution through growth alone can take much longer than what social consciousness allows" (Brasil, 1974, p. 52). Moreover, the 1974 document claimed that "the structure of income distribution is unsatisfactory, and with it the Revolution does not sympathise, since such a distribution resulted from the long-run evolution of the economy, not from recent factors" (ibid). This latter admission reveals the extent to which the government deflected responsibility from discretionary policy to "long run" economic factors.

5.2.4 The Persistence of the 1960s Paradigm Shift

Under the post-1974 Geisel government, wage readjustments were finally put in line with observed inflation rates (see Figure 3.24), as constraints on collective action were eased (IBGE, 2006). From 1979, after the violent strikes across 1978-79, wage adjustments were updated every 6 months, and for the first time since 1964, adjustments were freely bargained by employers and employees without government interference (Kane and Morisset, 1993). This made the economy highly vulnerable to further price rises. Indeed, the combination of the oil shocks, the domestic boom and growing scarcities of resources, as well as the inertial element in the incomes-indexation system of 1965, made inflation treble between 1973 and 1980, from 30% to over 90% (see Figure 3.1). With pressure to continuously raise average living standards and modernise the economy, growing inflation meant that inequality would again take a back seat in government plans. During the 1980s inflation became the number one priority, alongside growth (after the economy stalled). Inequality would thus be left to the whims of regressive inflationary pressures. Indeed, our results from the mid-1970s to the late-1980s shows a widening dispersion of income, as most of the growth was being captured by economic elites in a context of unequal inflation protection mechanisms (see section 4.2). Personal income inequality and redistribution would not become high-agenda topics until the late-1990s. However, socio-economic dimensions of inequality did gain much traction, especially in the 1988 Constitution and its aftermath. This focused on equalising rights and responsibilities between economic units, such as between federal, state and local governments, and between economic agents, such as between men and women.

[85] These potential dynamics are similar to the ones posited in chapter 2 in the context of the apparent fall in labour income inequality during the 2000s.

Our results in section 4.3 show how the gender question made much more progress than the income question in general.

After the stabilisation in the mid-1990s, and the renewed economic growth of the 2000s, inequality took a sharp upturn. Chapter 2 provides a detailed analysis of factors driving inequality over this period. We can add here that long-term institutional transformations preceding the period were necessary contributing factors, whose unequalising effects were only perceptible after the economy began to grow strongly again in the 2000s. Two of the most significant changes occurred in the areas of taxation and capital-ownership. As mentioned, capital-friendly tax polices were implemented from the late-1960s. These gradually removed capital incomes from the progressive income tax schedule. Rather than being subject to progressive marginal rates reaching between 50% and 60% for other top incomes until the late-1980s, a certain portion of capital incomes could be directly deducted from the tax base or could be exclusively taxed at lower flat rates (Nóbrega, 2014). The transition to democracy and universal suffrage in 1985–1988 actually pushed these regressive tendencies even further, following international trends in tax policy. First, with the slashing of top marginal tax rates on taxable incomes, from 50% in 1987 to 25% in 1989 – a rate that would be maintained until the present top rate of 27.5% was established in 1997.[86] And second, with the complete removal of distributed corporate profits and dividends from the personal income tax base starting in 1996.[87] Research, mainly focusing on the US case, has shown that large cuts to top income tax rates and the application of exclusive incentives to capital income can lead to the explosion of top incomes, through salaries and stock-related pay (Piketty et al., 2014, 2018; Lazonick, 2014).

The second significant policy change affected patterns of public and private ownership of enterprises. Similar to the changes in the income tax code, this change occurred in a context of shifting international trends in corporate governance and perceptions of state ownership. In Brazil, the scope of public ownership expanded notably in the post-war period as a necessary means to accelerate the process of industrialisation. This trend was consolidated further right up to the last years of the military regime.[88] During the economic chaos of the 1980s, privatisation tendencies began to take off. Between 1981 and 1989, 38 publicly-owned firms were sold, most of which were previously small/medium sised private firms who were taken over by the national investment bank (BNDES) (Baer, 2014). It was during the 1990s, however, when an explicit institutional turn towards privatisation emerged. In 1990 the new government passed the "National Privatisation Program", which established formal procedures for the selling of state-owned enterprises (SOEs). In the first half of the decade, most manufacturing SOEs were privatised, while in the second half

[86] The 1989 tax reform was directly influenced by the 1986 US tax reform (Silva et al., 2015).

[87] See section 1 of Appendix B

[88] According to data we exploit for the imputation of undistributed corporate profits into our estimates of inequality, of the 5,000 largest firms in 1974, 37% were public enterprises, while by 1985 the public share of the largest firms was 48% (see section 3.4.2 above, and section 6.2 of Appendix B).

privatisation accelerated further to include mining firms and public utilities – the latter also affected by the Concessions Law of 1995 and the constitutional amendments approved in 1996 later, which regulated concessions in public utilities.[89] In total, between October 1991 and December 2005, over 120 SOEs were sold (amounting to 87.8 billion US Dollars), with close to 70% of receipts coming from sale of federal SOEs (Baer, 2014).

Various features of this institutional process can be linked to the national income inequality trends we estimate for the post-stabilisation years: of the mentioned privatisations 64% of the sold capital went to domestic firms; a highest-bidder approach meant that most of the enterprises were taken over by the largest private domestic groups; while mergers and acquisitions (M&As) exploded over the decade (58 cases in 1992 to 351 in 1998), partly motivated by the need to pool revenues to make large successful bids for SOEs (Baer, 2014, p. 240-244). The de-nationalisation of enterprises was no doubt an important contributor to the observed increase in the private share of corporate assets and profits: according to the survey and national accounts data we use to impute undistributed profits to households (see section 3 and Appendix B), the private domestic share of corporate assets went from around 43% in 1985 to over 60% in recent years. Both the concentrated outcomes of the privatisation programs, as well as the penetration of international financialisation patterns in Brazil, can explain the positive effects on inequality since the 1990s. These trends are consistent with the striking evolution of billionaire wealth in Brazil, as depicted in Forbes'"The World's Billionaires" list. Between 1988 and 2014, the number of billionaires in the country increased from 3 to 65 (an 11.6% annual growth rate), while their total wealth increased by 12.1% per year (4 times as much as the growth in total national income over the period).[90]

Taken alongside our estimates, these two major policy transformations in Brazil illustrate that the link between re-democratisation and inequality is not a trivial one. In general, the economic and political changes of the 1980s and 1990s produced both concentrating and compressing effects in the distribution of income in subsequent years. On the one hand, the high-income subsidies channelled through the income tax reform and the privatisations of productive capital were accompanied by rising income concentration in the post-stabilisation period. On the other hand, "premature" de-industrialisation (promoted by the interchange of extreme inflation, high interest rates, trade and capital account liberalisation, overvalued exchange rates, and a fall in total investment), as well as a steady rise in government transfers and real minimum wages, gave way to an "unskilled-biased" technological-and-institutional-change paradigm, at least within the Bottom 99% of the distribution.

[89] These amendments discontinued public monopolies in telecommunications, distribution of gas, and in the oil sector. Rising utility rates followed these changes (Baer, 2014).

[90] See https://www.forbes.com/billionaires/list/.

5.2.5 Theoretical Perspectives Revisited

In section 2, we noted that the most pervasive theory linking growth and inequality was the redistribution of income to high-income-high-saving individuals, under the assumption that saving is automatically transmitted into productive investment to the benefit of the entire society. From the preceding sections, the views and actions of Brazilian statesmen and policymakers varied in their degree of consistency with this theory. During the period 1950–1964 there was a public purpose mission-oriented mentality with the "worker" as the protagonist, and the growth of formal employment and salaries as firm goals to expand the internal market. After the 1964 military *coup*, the economy became increasingly geared in favour of the high-income-high-saving class (or of the still scarce "highly-educated" and "productive" class, according to the official government line of the 1970s).

It is interesting that the economists linked to the military government, and government officials themselves, largely withheld from using wage-squeeze (or "forced-saving") arguments in the debates over growth and inequality patterns. As noted by Taylor et al. (1980, p. 327) their use could well have unleashed numerous "explosive social issues". As the authors' state: "why should workers have to pay through lagging nominal wages for the investment of the rich or of the state [or of foreign enterprises]? In the models just described, different savings patterns among classes may explain *why* workers pay under current government policy, but they scarcely justify the existing arrangements. In fact, under socioeconomic policies, ranging from a more humane tax structure up to outright expropriation, growth could be maintained or accelerated without making the workers worse off" (ibid).

We would argue that similarly sensitive social issues were at the core of the 1960s paradigm-shifting crisis. The intentions to overcome structural bottlenecks to further industrial development – associated to outdated rural-property, electoral and tax regimes and profit reinvestment – produced distributional bottlenecks, as they were met by stern political opposition (both at home and abroad). This only intensified with the economic deceleration after 1962, which the government exacerbated by cutting spending and applying tight monetary policy, while it maintained its wage policy in favour of salaried earners, at the expense of top incomes. The ensuing economic and political crisis produced radical uncertainly, halted investment plans, and thus eliminated the autonomous components of demand that could have lessened the economic decline (Serra and Tavares, 2000). This panorama bears some resemblance to the crisis and paradigm shift that engulfed Brazil over 2014–2018, which began with a recession and culminated with Dilma Rousseff's impeachment in 2016. Distributive reforms were equally delayed, but economic elites were not as squeezed as they were by the early 1960s, as our results show.

The repressive response to the 1960s crisis is consistent with Hirschman and Rothschild (1973) "tunnel effect" hypothesis for segmented societies. Indeed, one could argue that the belief of the new government of the early 1960s that it could simultaneously achieve

(or had a sufficiently ample mandate to simultaneously pursue) higher growth *and* equity in the income distribution was over-estimated, given the highly segmented structure of society – regionally, racially and politically.[91] In this respect, it is important to note the absence of large-scale proximate wars or of landmark revolutions in Brazil, which could have unified the population under a common cause, as happened in the USA, Western Europe, China and Russia to varying extents. For instance, Brazil was the only country of the region whose colonial independence was proclaimed by members of the colonial class rather than fought from them.

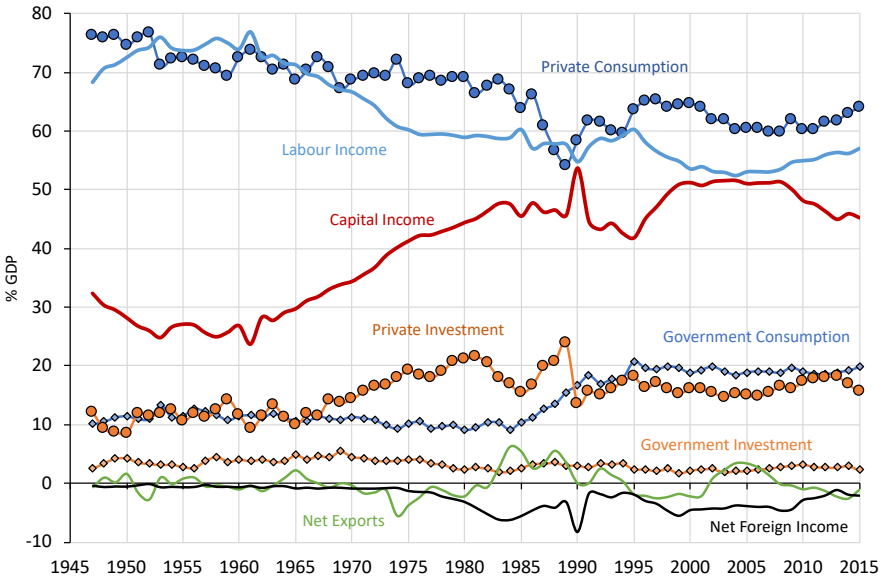
Another way to assess the relevance of the wage-squeeze-forced-savings (WS-FS) theory in Brazil is to relate our inequality estimates with macroeconomic data to evaluate its predictions. These state that the redistribution of income to high-earners stimulates growth as the extra surplus will be invested in expanding productive capabilities. During the military dictatorship in general this pattern seemed to be present. From the late 1960s private investment plans began to pick up after clear signs and trends that the economy would be engineered in favour of capital income (see Figure 3.26). This would only increase over the strong growth of the "miracle" years of the 1970s, with private investment growing at 13% per year up to 1980, considerably more than any other GDP component. At the same time, government investment retracted as a share of the economy, while that of foreign capital investment rose, especially after 1974 (this can be seen from the marked rise in net foreign income in Figure 3.26, practically all of which is related to foreign capital income, as observed in the national accounts). Thus, the data seem to show a strong correlation between the increase in the capital-wage ratio and investment-led growth during this period.

Yet, such a co-variation does not necessarily settle the question in favour of the WS-FS theory to guarantee this outcome. Indeed, two counter-examples relate to the 1950s and the latter-2000s. In the former period, labour income was growing faster than capital income, and government investment underwent its faster growth of any period (driven by the 17% annual expansion from 1956–1962). Private investment grew practically as much over this period as it did over the 1967-1980 period (12% vs 13% per year). Similarly, from 2007 to 2013, wages rose faster than capital income and strong growth in public investment (from Lula's PAC program), at 4.1% per year, was equally matched by the upswing in private investment, at 5.4% per year. However, despite the renewed investment growth in recent years, the share of capitalist investment in total capital income has not yet recovered enough to reach the levels of the late-1950s/early-1960s or the late-1970s/early-1980s (see Figure 3.27). Since the return to democracy capitalist consumption and saving have outpaced capitalist investment.[92]

[91] Only 40% of the voting age population were participating in elections in the early 1960s, due in part to the literacy requirement on voting, which nominally excluded approximately 40% of the eligible population (IBGE, 2017) and `http://www.ipeadata.gov.br/Default.aspx`.

[92] See chapter 2 for an analysis of patterns of corporate investment and saving during the 2000s. Between 2010 and 2015 investment of non-financial firms receded by 13%, while their net acquisition

Figure 3.26: Income and Expenditure Components of GDP: 1947–2015



Notes: GDP components are from IBGE (2006, 2017). Labour and capital shares correspond to gross shares in GDP, which are equal to the gross incomes in Figure 3.22 minus the net foreign income associated to each factor. Government consumption expenditures exclude monetary social transfers and benefits, which are included in private consumption. These have increased from approximately 8% of GDP in 1988 to 14% in 2015 (Morgan, 2018).

What this teaches us is that the WS-FS theory may provide *sufficient* conditions for investment-led growth, but it does not provide the *necessary* conditions. Both during the 1950s and 2000s private and public investment simultaneously increased their share in GDP, at the same time as capital income was falling. These were periods when real minimum wages were rising strongly and when government investment was a leading stimulant to the economy. The ultimate distributional difference between the 1950s and the 2000s comes from the productivity-wage gap sourced from late-1960s wage repression (see Figure 3.23), making the legacy of the military dictatorship and its growth strategy still visible today. In this respect, the sufficient conditions of the WS-FS theory are only likely to be applicable during a dictatorship. Thus, coercion becomes a necessary condition for the WS-FS policy to sustain investment-led growth over the long-run in a developing economy. This is a point not commonly made in the literature on inequality and development.[93]

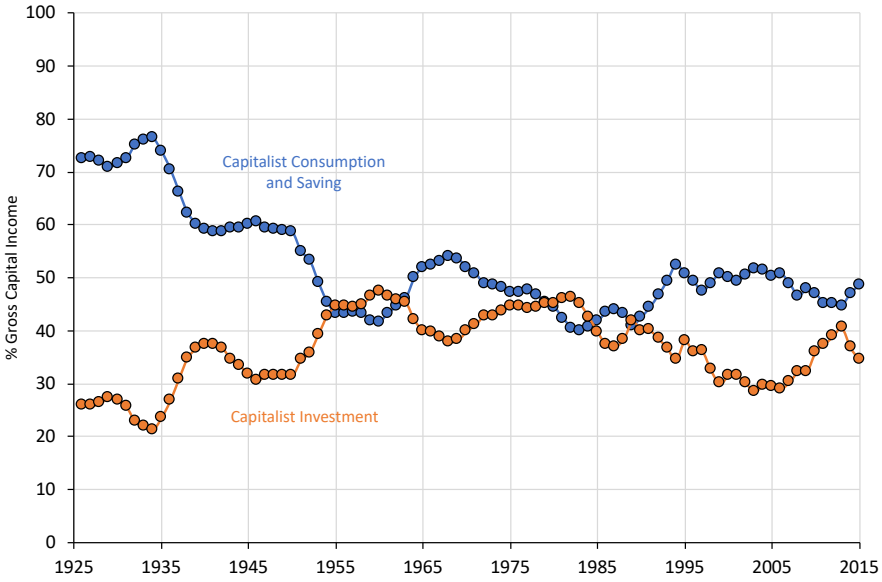
In the 2000s, minimum and median incomes increased more than average income, but the Top 1% average income grew faster than all three; while during the 1950s, minimum and median incomes were growing faster than both average income for the full population *and* the average income of the Top 1%.[94] What our analysis also reveals is that the

of financial assets rose by 40%. This is consistent with Figure 3.27.

[93] Hirschmann is one exception.

[94] Unfortunately, for the earlier period we do not currently know how income was distributed

Figure 3.27: Capitalist Investment vs. Consumption and Saving: 1926–2015



Notes: the shares of capitalist investment (private gross fixed capital formation) and capitalist consumption and saving (net of foreign capital outflows and capital income/property taxes). Capital consumption includes consumption of fixed capital. Authors' computations from data in IBGE (2006, 2017). We take 5-year moving averages to smooth the series for 1947–1995, and annual estimates from 1926–2015.

necessary conditions for equitable and sustainable development hinge on the political economy related to the country's social and political structure. This, we argue, was the key element determining the lifespan of the post-war "progressive" regime in Brazil.

In sum, Brazil's distributive history shows the clear importance of institutions in connection with the economy's development lag. The theorists that explicitly deal with social conventions and social structure (Marx, Keynesian theorists, Hirschman) and underdeveloped peripheral countries, alongside their dynamic interaction with an advanced centre concerning trade, investment and institutions (developmentalist–structuralist), are of particular relevance here (see section 2). The Brazilian case also illustrates the need to go beyond simple classical-dualist theories of development, such as those associated to Lewis (1954) and Kuznets (1955), largely because they overly abstract from the political cycle and social structures. While Kuznets recognised the importance of institutional variables for the downward cycle of his growth–inequality nexus, he does not make the link between industrialisation and inevitable distributional bottlenecks that arise in countries engaging in late-development. This is largely due to his focus on more the more developed regions at the time. The relevance of neoclassical theories is even more challenged, to the point of complete irrelevance. As we have shown for Brazil, these bottlenecks are underpinned

among the Bottom 99%.

by social conflict over the direction of institutional transformation necessary for further development. This is consistent with endogenous theories of institutional change, such as the one developed by Blyth (2002).

The political economy of the 1950s and early 1960s in Brazil meant that Lewis' "reserve army" of traditional sector workers could not keep industrial wages at a subsistence level. All our indicators show that wages in the modern sector rose slightly more than productivity, as successive governments were eager to empower their most important electoral cohort. But we cannot confirm Kuznet's and Lewis' hypothesis that inequality (concentration) was higher in this sector than in the primary sector. We can state that the distance between top incomes and other incomes was falling. The paradigm-shifting *coup* of 1964 replaced this political settlement by one where economic elites could absorb a higher surplus and income share, while workers in the politically important agricultural sector/less industrialised states could capture a higher share of the productivity gains. History has not exactly repeated itself (yet) in the 2014–2018 period, but similar tendencies have been in the making.[95] The reasons the 2000s did not replicate the 1952-1964 distributional dilemma to the same degree are precisely due to the institutional legacy of the dictatorship and post-dictatorship periods for the economy. Given, that growth and development were curtailed after the 1980s, distributional pressures had a greater distance to travel.

5.2.6 Development, Inequality and Inflation

Throughout the entire period of our analysis inflation was a hotly disputed issue in Brazil - as important as inequality itself. As described in previous sections rising costs of living fed larger and larger increases in nominal wages, when "labour" governments and strong worker mobilisation were present. This settlement arguably over-extended its reach, in producing the "stagflation" of the early 1960s. Although the military dictatorship managed to quell inflation through wage and monetary repression initially, a combination of external supply shocks, the economic boom, internal pressures on real resources, and the incomes-indexation policy quickly led to mounting price spirals that would come to define the 1980s and early 1990s.

For these reasons, we think it necessary for distributional analysis, especially in late-developing countries, to incorporate a dynamic theory of development, inequality and inflation. A general conjecture can be made: the combination of high inequality and low living standards, during the pursuit of development, is one of the root causes of persistently high inflation. In Brazil this has been due to the very failure to adequately redistribute the "fundamentals" (land, capital, income and education), which puts pressure on wage inflation and material limits on the development that can be attained in "democratic" times.

In Brazil, post-war governments (especially those from 1951), in which labour unions

[95] See chapter 2 and Gethin and Morgan (2018).

were strongly represented, attempted to address the problems of late-development and inequality simultaneously which produced greater price rises. Anchoring periodic wage adjustments to changes in the cost of living accelerated inflation. Likewise, during the wage-easing years of the dictatorship (post-1974), the policy of indexation meant that the periods between successive wage revisions became ever and ever shorter because of the very nature of the adjustment process – adjustments were mainly based on the accumulated change in living costs relative to a previous benchmark date. This cumulative inertial process to inflation was predicted by Kaldor (1957) in his consideration of inflation in underdeveloped countries. This process was further exacerbated by subsidised government financing of firms (Serra and Tavares, 2000). Indeed, as highlighted by Serra and Tavares (2000) the inflation of the 1950s "artifically" maintained the profitability of firms, especially in sectors linked to strong investment in physical capital of the late 1950s. Thus, it temporarily mitigated "the profit-wage tensions by preserving "illusory profit rates" for new investments." Capacity constraints, concluding rounds of federal mission investments, as well rampant government wage increases eventually diminished the marginal productivity of capital. Both in the early 1960s and in the early 1980s, inflationary pressures were eventually counter-acted by tight monetary policy and temporary fiscal austerity policies, which produced a fall in investment and economic contractions, without structurally ending inflation. This would only be reached after 1995, when the *Plano Real* changed expectations, and when wages and the economy in general had receded enough. As shown in Figures 3.1 and 3.2, Brazil has since been playing catch-up.

Macro imbalances culminating from rising inflation and distributional conflict have often led to major disruptions, resulting in military coups, presidential impeachments, etc., and a continuously revolving social order. As a result of the un-addressed distributional imbalances, subsequent governments in Brazil have been compelled to spend, not only to foment development, but also (through social spending) to keep society "at peace" (note the significant rise in government consumption after the return to democracy in Figure 3.26).[96] At the same time, high interest rates have become a persistent structural feature of the economy given past experiences with inflation. And so too has inequality.[97]

6 Concluding Remarks

In this paper, we provide new evidence on inequality over the long-run of a late-developing country. We are the first to reconcile micro and macro distributions of income for Brazil, using recently developed accounting frameworks (Alvaredo et al., 2017) and statistical

[96]Since 1985 the general government has run a primary deficit of around 6% of GDP on average, compared to the 1% surplus run between 1964 and 1985 the 3% deficit between 1945 and 1964. Figures are computed from data in IBGE (2006) and https://www.tesouro.fazenda.gov.br/balanco-do-setor-publico-nacional-bspn-.

[97]See chapter 2 for more details on the recent period.

techniques (see chapter 1). The Brazilian case is a interesting one, particularly for the lessons one can draw from it.

Our results reveal the huge magnitude of income concentration and dispersion among the Brazilian population across ten tumultuous decades. The experience of most other countries pales in comparison to Brazil's inequality history, although some, more than others, seem to be converging to the Brazilian frontier, both at the top and at the bottom. The Top 1% national income share rarely deviated outside the 20-30% bandwidth – when it has, it has more often that not been beyond the upper threshold. We find that Brazil's inequality, particularly its income concentration, is tightly correlated to the country's marked shifts from one political regime to another.

We show that economic elites within the top percentile have appropriated higher income growth on average since the 1920s. But the variation across different eras shows the extent to which different institutional paradigms determined how growth was distributed among the different strata of the population. The resilience of the Top 1%, particularly in maintaining the share of the top decile stable since the 1970s (while the Top 10-1% share continuously fell), highlights how the top has successfully relocated the distributional conflict lower down the distribution – as one effectively between the poor and the "middle class". Contrary to expectation, this pattern has only been exacerbated further since re-democratisation and universal suffrage. The role of the early 1960s crisis, the subsequent legacy of the military dictatorship, as well as the embrace of international patterns of economic policy following its end, are crucial events to understand the persistent prosperity of elites in Brazil over recent decades. Related to these dynamics, we reveal the increasing irrelevance of surveys to adequately shed light on income disparities across the distribution. To this end, we motivate the need for data harmonisation procedures in the manner we have exposed in this paper.

Overall, the Brazilian economy experienced rapid change in structure and size over our period of analysis. Yet this was not fully matched by institutional changes in a country historically fraught with acute social divisions. Elites have been successful in managing to keep these divisions at bay, especially when they have compelled policy changes. The absence of one-off elite-compromising shocks (wars, revolutions) or sustained social-democratic mobilisations like those other countries experienced broadly explain why Brazilian inequality has consistently remained comparatively high. However, contrary to Williamson's (2015) hypothesis, Brazil did not entirely miss the"20^{th} century egalitarian leveling", evidenced by the continuous reduction of top shares between 1942 and 1964.

We pinpoint two episodes of strong inequality reduction. The first relates to the 1950s, when strong industrialisation and urbanisation were accompanied by rising urban wages more than in line with productivity. This arguably had gone too far by the early 1960s, when inflationary wage policy continued in the presence of the first agricultural labour laws and national worker confederations. But its unsustainability was *especially* so given the unresolved imbalances regarding the access to capital (land, business capital)

and education (determining labour market insertion *and* electoral participation). With government executives seemingly un-attracted by the state-communist solution (despite suspicions of conservative-military factions of the link between government and the growing syndicalist movement), something had to give way: the fall of either the government or of elites. Society succumbed to the former by force in April of 1964. Five decades and multiple crises later, the election of another "labour" government could not fundamentally alter the inherited distributive paradigm, despite achieving notable improvements in material living standards of the poor.

The second progressive evolution we find relates to the rise of women in the labour market, which in itself is a remarkable finding. This is especially true from an international comparative perspective. In 40 years the average gender-wage gap in Brazil was reduced by more the 60%, as the share of women in the labour force jumped from just over 25% in the mid-1970s to just under 45% today. At the same time, the representation of women among top earners grew even faster, with the share of women in the Top 1% increasing by a factor of 7 between 1974 and 2016. This growth has resulted in Brazilian women out-representing their counterparts in developed countries like France and the USA at the top of the earnings distribution. If the concentration of income growth was overly resilient, it seems that socio-cultural dimensions of inequality were less so. We will investigate this further in future work.

***

Reading what Brazilian government officials and intellectuals wrote and declared, all pretty much recognised (albeit to different degrees) that living standards were unacceptably low and inequality (the "social question") was an issue. While growth was always seen as the answer to avoid social conflict in advanced western countries, in a country like Brazil – poor and highly unequal – it was perceived to be more urgent, to the extent that "development", rather than "growth" *per se*, was the panacea. The idea of most conservative thinkers (especially associated to post-1964 governments) was that Brazil would grow rich *and* more equal over time. While this resonates with (Neo-)Classical-Dualist theories of inequality and development, it underestimates the extent to which development, through industrialisation, engenders distributional bottlenecks that demand deep-seated institutional change. Thus, theories that don't abstract from institutional details (Marxist-Keynesian-Structuralist) are far more relevant to understand the case of late-developing countries like Brazil.

Indeed, in the context of the intense debate on the 1960s stagnation and distributional shift, our results and analysis reveal that it had more to do with institutional changes (in anticipation of other impending changes) than with market-based human capital explanations, as the initial dissidents (Hoffmann, Duarte and Fishlow) measured. The prolonging of these distributional imbalances after 1964 also reveals the extent to which coercion may itself by a necessary condition to ensure that a wage-squeeze-forced-savings (WS-FS) policy gets passed on into investment-led growth – a conclusion not commonly

made in the theoretical literature on the subject. Brazilian experience too shows that development outside of a WS-FS paradigm is possible, but, as argued, it will inevitably face distributional constraints that require an institutional revamp. And this change, as well as the resulting inequality dynamics, can go in either direction, depending on the balance of power of different social actors.

The balance between democracy and dictatorship in such an overall context hinges on the "social question". Brazil knows this too well. The military dictatorship could implement a very regressive austerity policy precisely because it was a dictatorship. But even this regime was eventually relented, exemplified by its second national development plan, from 1975. Distributional imbalances were sidelined with the inflationary and growth crisis, which overshadowed the initial stages of re-democratisation and universal suffrage. But since then, the Brazilian economy has been forced to keep growing to accommodate all social demands, given that the "fundamentals" continue to be neglected. But as history keeps revealing, this can only be prolonged up to a point. In Brazil, this point has historically been located at the junction between development, inequality, and inflation. To paraphrase Hirschman, only further development will tell where the future of the current fragile democratic settlement in Brazil lies.

Appendix A

The Weight of the Rich: Appendix

Abstract: This Appendix presents more details on the data sources, results and analysis of chapter 1.

1 Country Specific Income Concepts and Observational Units

1.1 Brazil

In reconciling incomes in surveys with those in tax data, we use the latter as the benchmark for the top of the distribution. We thus require that the survey definition of income, from the micro-data, be consistent with the definition of income in the tax tabulations in order for the comparison to make sense. The total income assessed in tax data is pre-tax-and-transfer income, but including pensions and unemployment insurance. It is the sum of three broad fiscal categories: taxable income, exclusively taxed income and tax-exempt income (reported in Table 9 of the tax report *Grandes Números DIRPF*). We describe each of these in turn before describing how we construct the survey definition of income.

Taxable income comprises of wages, salaries, pensions and property rent. These are incomes that are subject to assessment for the personal income tax. Exclusively-taxed income is income that has been already been taxed at source according to a separate tax schedule. It also contains capital income and labour income components. The labour component is the sum of the 13th monthly salary received by the contributor and their dependents, wages received cumulatively by contributors or dependents, and worker participation in company profits. The capital component comprises of the sum of fixed income investment income, interests on own capital ("juros sobre capital próprio"), variable income investment income, capital gains and other capital income. Non-taxable incomes are the last fiscal category, whose decomposition is presented in Table 20 of the tax reports. These are incomes that are declared but which are not subject to any personal taxation when received. Close to one-fifth of these exempt incomes can be classified as labour income. These comprise of compensation for laid-off workers, the exempt portion of pension income for over 65s, withdrawals from employment security fund, scholarships, and other labour incomes. The remaining items can be classified as capital income (distributed company profits, dividends, interests from savings accounts/mortgage notes) or mixed income (the exempt portion of agricultural income).

We construct survey income to be as close to the tax definition as possible. The total income we analyse from the PNAD surveys is the sum of labour income, mixed income and capital income. Labour income is the sum of all reported income from primary, secondary or all other jobs (variables V9532, V9982, V1022) for all employed individuals who do not classify themselves as own-account (self-employed) workers or employers. For employers, we assume that labour income is the portion of their work income that is below the annual exemption limit for the DIRPF, as set by the Receita Federal. Thus, values above the first tax paying threshold are taken to be capital withdrawals. Also in labour income are pensions (V1252, V1255, V1258, V1261), work allowances (V1264) and unemployment

insurance. The latter is taken from other income sources declared (V1273) and estimated as income from this source that is reported between 1 and 2 monthly minimum wages. Values of V1273 equal to or below 1 monthly minimum wage are interpreted as social benefits, which are excluded from the analysis.

Mixed income is the reported income of own-account workers. Capital income is estimated as the sum of rent (V1267), financial income, and the capital portion of employer work income (i.e. reported amounts exceeding the annual exemption limit for DIRPF). Financial income (interests and dividends) is taken from other income sources declared (V1273) and estimated as any income from this source that exceeds 2 monthly minimum wages. Finally, we add a 13th monthly salary to the annual calculation of the incomes of formal employees and retirees. In total, the income we calculate from the surveys represents close to 80% of the equivalent (fiscal income) total from the household sector in the national accounts, on average between 2007 and 2015. The total income we use from tax statistics accounts for about 63% of the same fiscal income total from the national accounts over the same period.

Given that the unit of assessment in the tax data can either be the individual or the couple, in cases where the latter opt to declare jointly, we cannot strictly restrict ourselves to the analysis of individual income as it is received by each person. Therefore, we decide follow the tax legislation by identifying the number of married couples appearing jointly on the declaration and splitting their total declared income equally between them when carrying out the generalized Pareto interpolation from the tabulation. This allows us to bring the analysis to the individual level by assuming that all spouses equally share their income. We use the information available in the tax statistics to estimate the share of joint declarations, which overall represent about 30% of all filed declarations (see 2). To be consistent in the comparison, we also use individual income in the surveys, with the income of married couples being split equally between the composite adults. We consider all adults aged 20 or over in our analysis.

1.2 Chile

Following the same logic as that applied to the Brazilian case, we construct from the Chilean survey an income definition that is as close as possible to the one used in tax data. The resulting definition is the one we use when merging datasets. However, in Chile, unlike Brazil, the survey reports post-tax incomes. In broad terms, we estimate pre-tax income retrospectively from declared post-tax income. In order to do so, we make a priori assumptions on whether certain types of income pay income taxes or not. Additionally, some self-reported characteristics are used to determine if the income of certain individuals should be treated as taxable or not. For instance, dependent workers that do not have a contract (and will not sign any soon) are considered to be informal, thus they are assumed to not pay the income tax. A similar mechanism is used for independent workers – depending

Table A.1: From Post-Tax to Pre-Tax Income in Chilean Surveys

Type of income	Taxable Income		Tax Exempt Income	
	Variable name	**Code**	**Variable name**	**Code**
Labor Income	Wage (1ry occup.).	y1a	Occasional work.	y16a
	Wage (2ry occup.).	y6, y10	Unemp. insurance.	y14c
	Inc. from previous months (if dependent).	y14b	Tips, travel expenses.	y3c, y3e
	Extra hours, commissions & allowances.	y3a, y3b, y3d y3f	Christmas bonus.	y4a
	Rewards & additional salary.	y4b, y4c, y4d	Inc. of the inactive.	y11a
			Wage of informals.	o17, o14
Pensions	Old age pension.	y27am		
	Disability pension.	y27bm		
	Widow's pension.	y27cm		
	Orphan's pension.	y27dm		
Mixed Income	Inc. of indep. (1ry occup.)	y7a	Inc. of indep. (2ry occup.).	y6,y10
	Inc. from previous months (if indep.).	y14b	Inc. of non-qualified, informal, small minery & craftsmen.	oficio1, oficio4, o14
Capital Income	Rent (agricultural).	y12b	Rent (urban).	y12a
	Interest.	y15a	Rent (seasonal).	y16b
	Dividends.	y15b		
	Withdrawals.	y15c		
	Rent (equipment).	y16a		

Notes: Codes correspond to those of CASEN 2011-2013. Formality is defined as conditional to having a contract and/or emitting "*boletas de honorarios*" (invoices by independents). Information on formality is only available for primary occupation. Formality is assumed to be the same for 1ry and 2ry occupations. In the survey, income is post-tax. Pre-tax formal income of contract-workers is calculated using tables of IUSC (*Impuesto Único de Segunda Categoría*) retrospectively. Pre-tax income of formals emitting invoices is added of mandatory provisional deductions (e.g. 10%) and standard presumptive expenses (e.g. 30%). Pre-tax capital income is calculated using the IPC (*Impuesto de Primera Categoría*) single tax-rate (e.g. 20%). Rent of urban properties is assumed to be untaxed because of law D.F.L.2 (1959)

on if they emit invoices (both commercial or for services) we define them as formal or informal. Table A.1 gives a comprehensive view on what types of income are assumed to pay taxes or not. For further comments on the definition of income corresponding to tax data, please refer to Atria et al. (2018).

1.3 European Countries

Tax Data For the three European countries we use tabulated tax data from official sources. In the case of Norway and the United Kingdom, the data come directly from institutional sources, Tax Statistics for Personal Taxpayers from Statistics Norway (`https://www.ssb.no/en/statbank/list/selvangivelse`) for the former, and the Survey of Personal Incomes (SPI) from HM Revenue & Customs (`https://www.gov.uk/government/statistics/income-tax-liabilities-by-income-range`), for the latter. The tax unit for both countries is the individual. As explained in Section 4.2, we interpolate the tabulations using a generalized Parteo interpolation (*gpinter*). For France, we use detailed tabulations produced by Garbinti et al. (2016) from the micro-files of French taxpayers. These are available in the Appendix C Tables of their Data. We use the individual-level tabulations that present the distribution of gross total fiscal income for 127 percentiles.

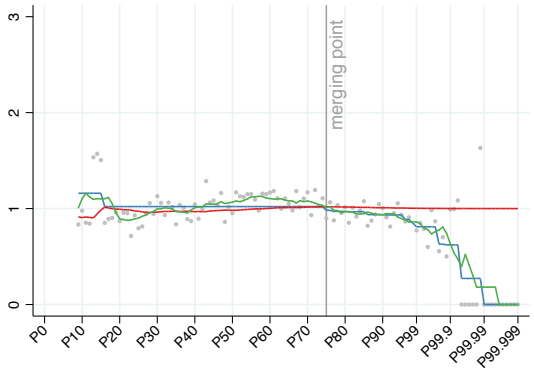
EU-SILC Data The advantage of using EU-SILC data is that it is a harmonized household survey dataset for European countries. However, given that we anchor our estimation method to the tax data, the definition of income used from surveys must match that accounted for in tax statistics. To do so we take the sum for each observation of employee cash or near cash income (variable PY010), self-employment cash income (PY050), Pensions received from private plans (PY080), a host of benefits related to unemployment, old-age, suvivors, sickness and disability (PY090, PY100, PY110, PY120, PY130), and capital income components (rent from property or land (HY040) and interests, dividends, profit from capital investments (HY090)). These capital incomes are reported at the household level. We individualise them by equally splitting the income among spouses and civil partners. For Norway and the UK, consistent with the fiscal income in tax data, we take gross incomes (before income taxes and individual social contributions levied at source). Since fiscal income in the French tax data is before income tax but after social contributions levied at source, we take net income values from the French SILC dataset. Income taxes are not levied at source in France for the period we analyse so the definition of net income in SILC is apt to be used for this case. We also select the reference population to be kept in accordance with the tax statistics. In Norway, the tax tabulations refer to individuals aged 17 and over, so we discard individuals under the age of 17 in the survey. For the UK, the tax data does not provide comparable information, so we follow the practice by Atkinson (2007) in taking a reference population of individuals aged 15 and over. In France, consistent with the use of the population aged 20 and over in Garbinti

et al. (2016), we keep persons aged 20 and over in the survey.

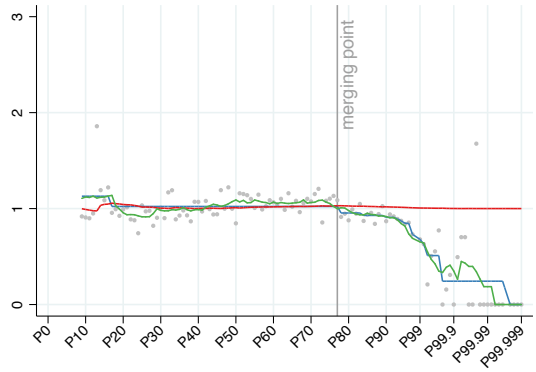
2 Shape of the Bias

Figures A.1-A.5 show the shape of the bias we estimate for the other years among our sampled countries. Each coverage of the data points are determined by the trustable span of the tax data in each country, which is defined as the portion of the population that are subject to positive income tax payments.

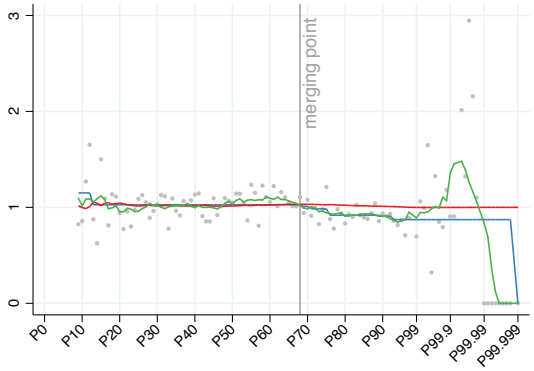
Figure A.1: Merging Points in Norway, 2004-2013



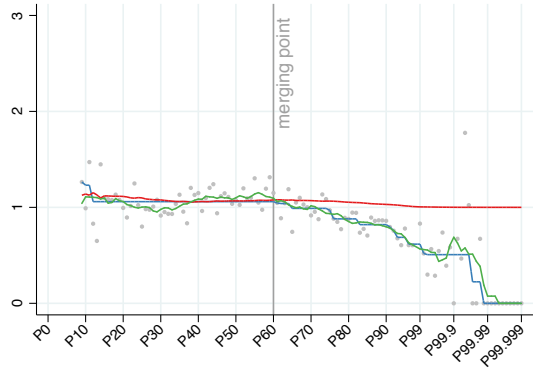
(a) Norway 2004



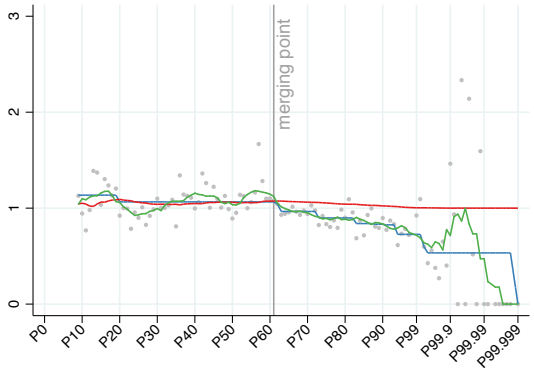
(b) Norway 2005



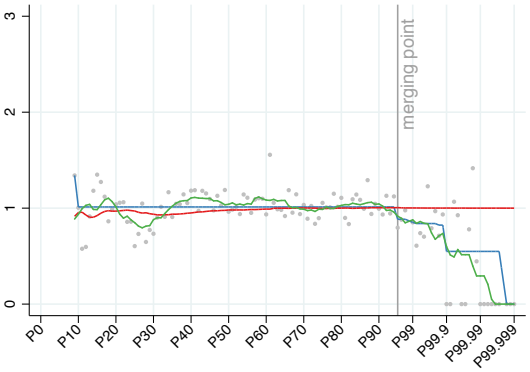
(c) Norway 2006



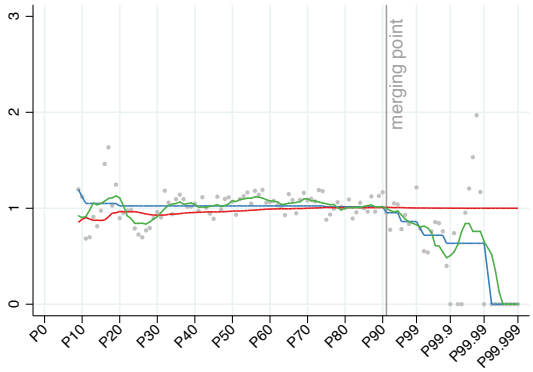
(d) Norway 2007



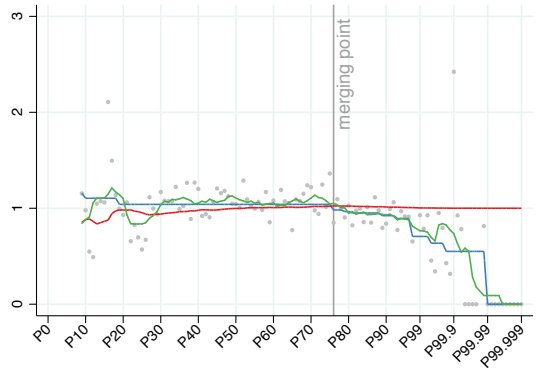
(e) Norway 2008



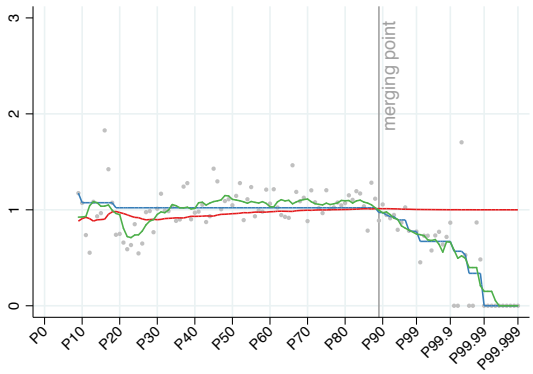
(f) Norway 2009



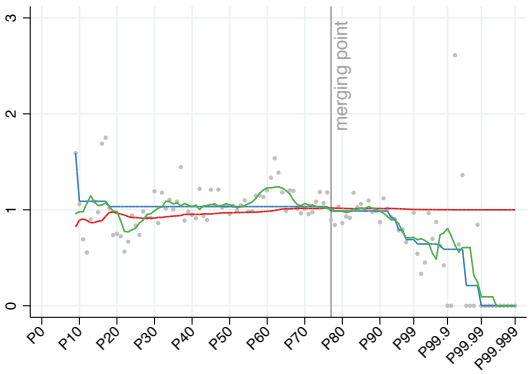
(g) Norway 2010



(h) Norway 2011



(i) Norway 2012



(j) Norway 2013

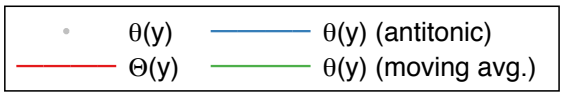
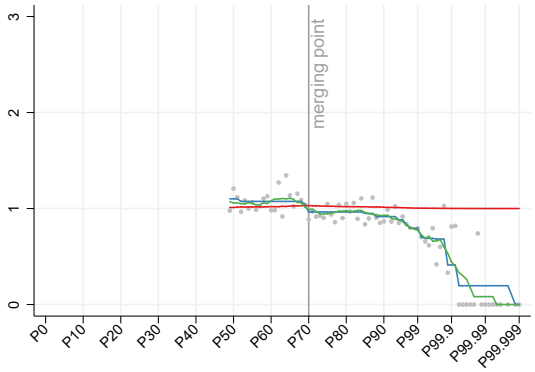
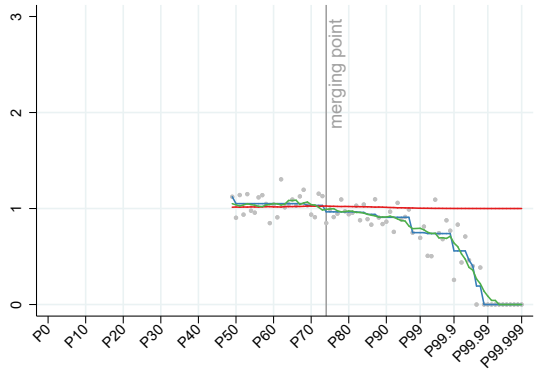


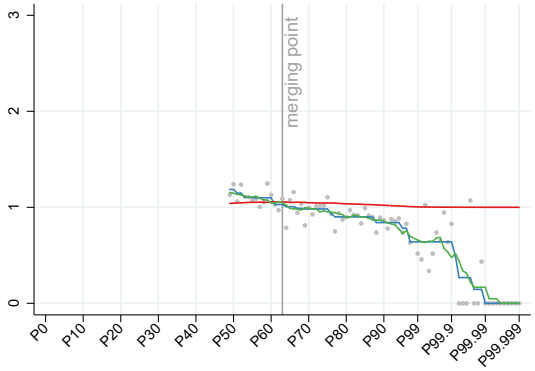
Figure A.2: Merging Points in France, 2004-2013



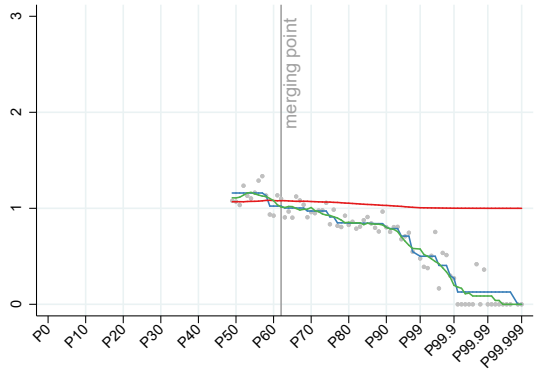
(a) France 2004



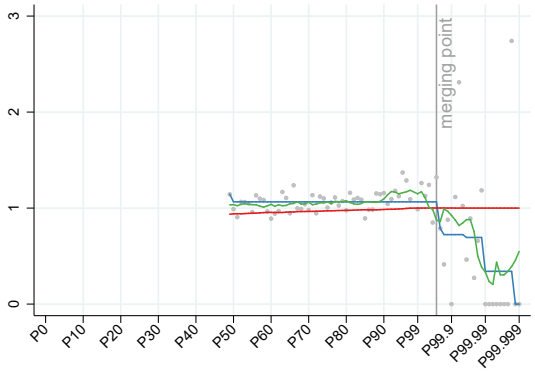
(b) France 2005



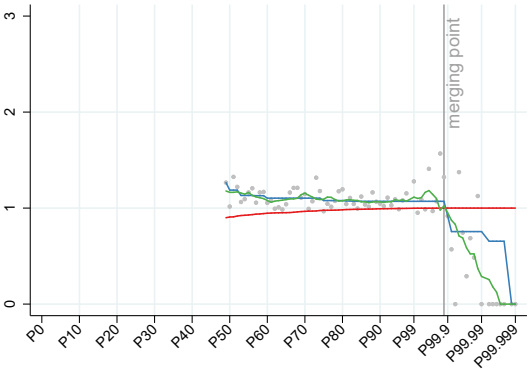
(c) France 2006



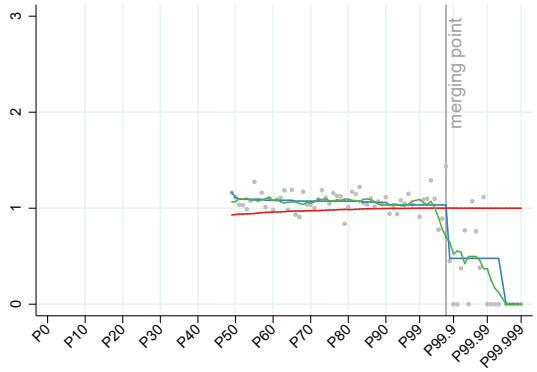
(d) France 2007



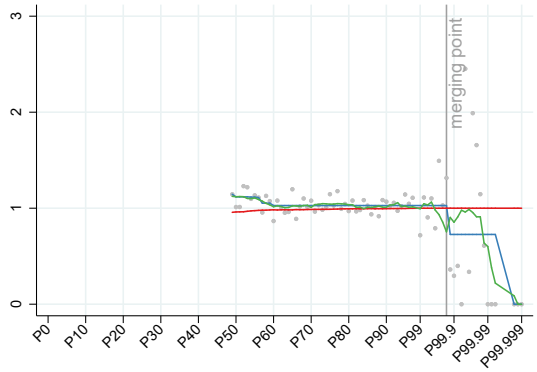
(e) France 2008



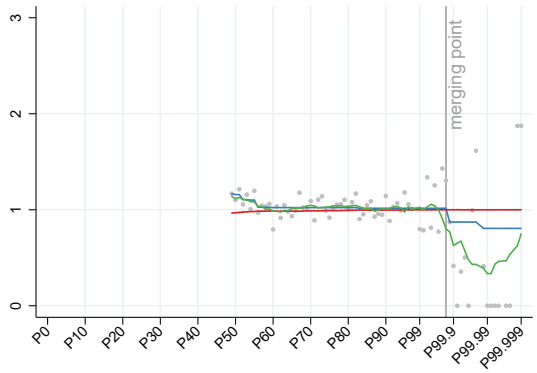
(f) France 2009



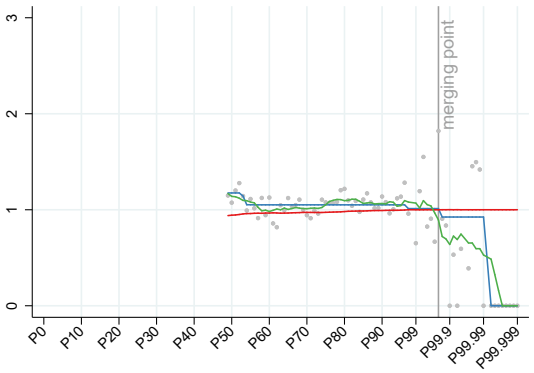
(g) France 2010



(h) France 2011



(i) France 2012



(j) France 2013

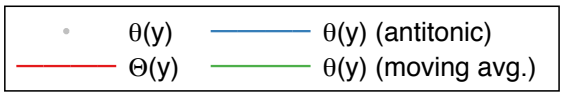
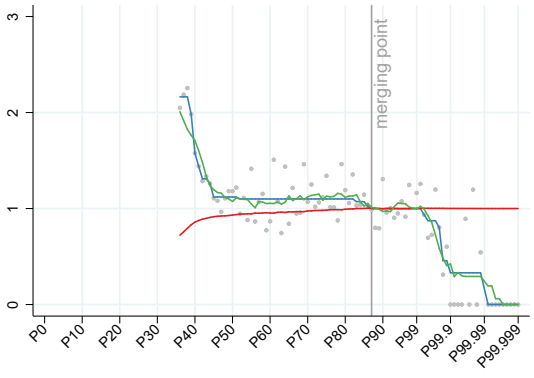
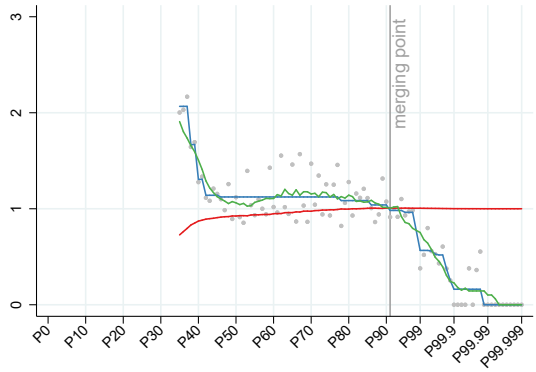


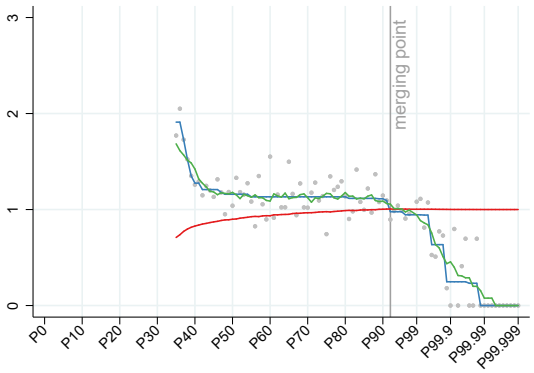
Figure A.3: Merging Points in United Kingdom, 2005-2013



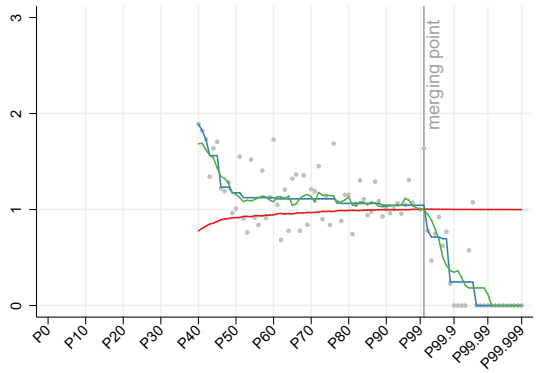
(a) United Kingdom 2005



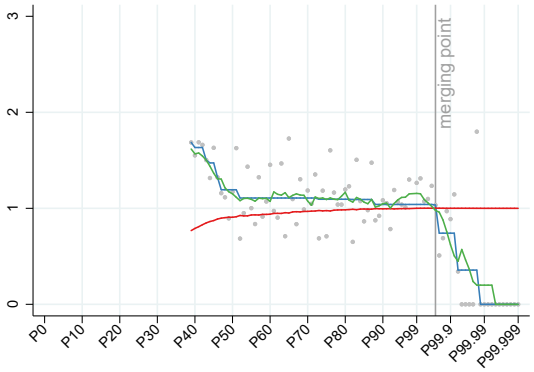
(b) United Kingdom 2006



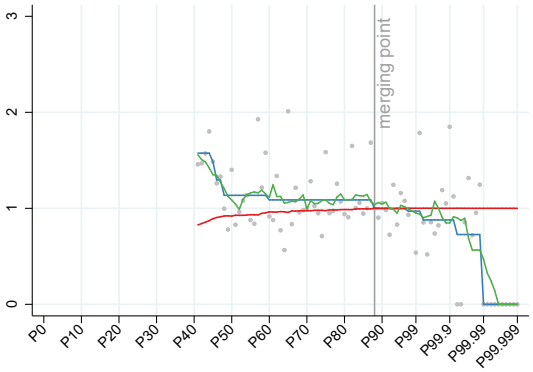
(c) United Kingdom 2007



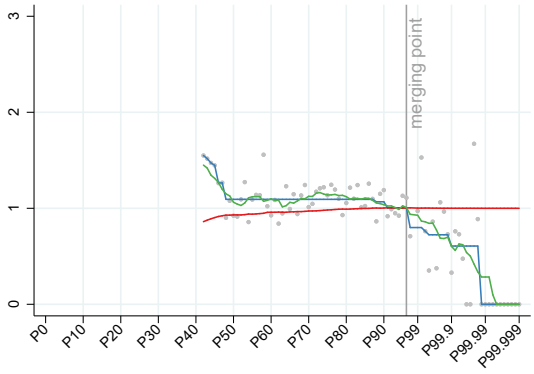
(d) United Kingdom 2009



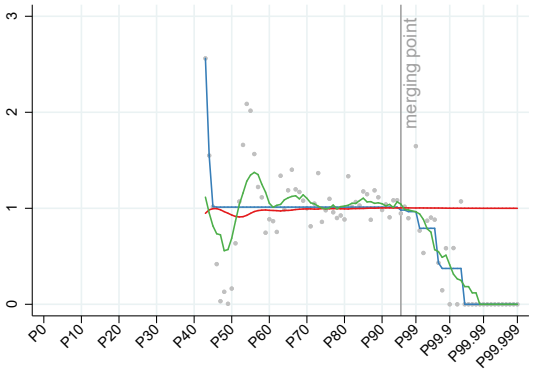
(e) United Kingdom 2010



(f) United Kingdom 2011



(g) United Kingdom 2012



(h) United Kingdom 2013

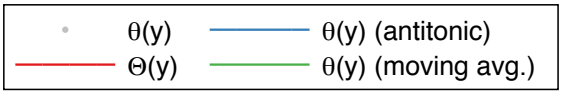
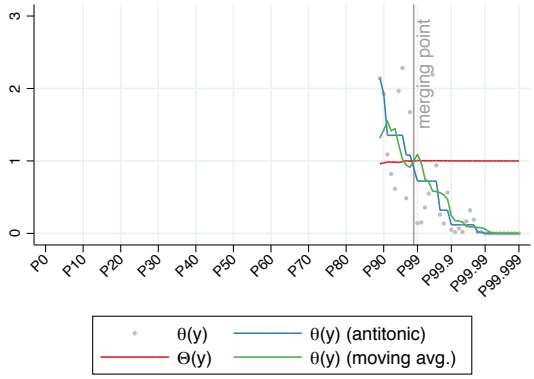
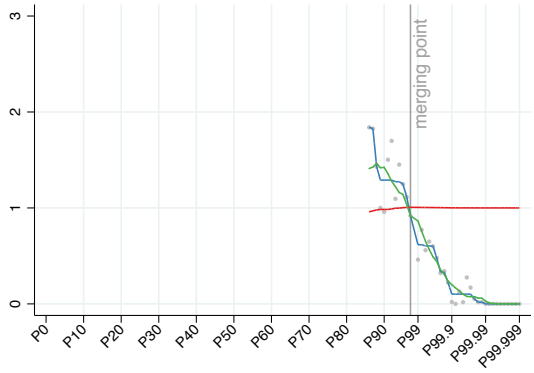


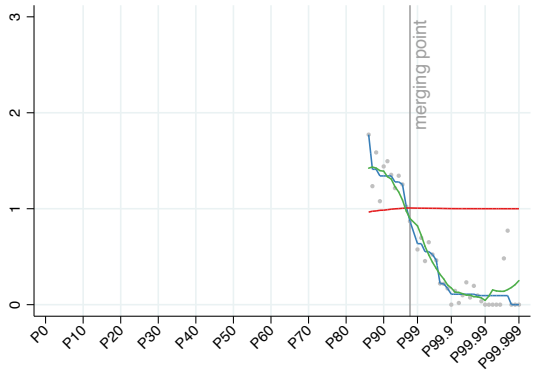
Figure A.4: Merging Points in Brazil, 2007-2014



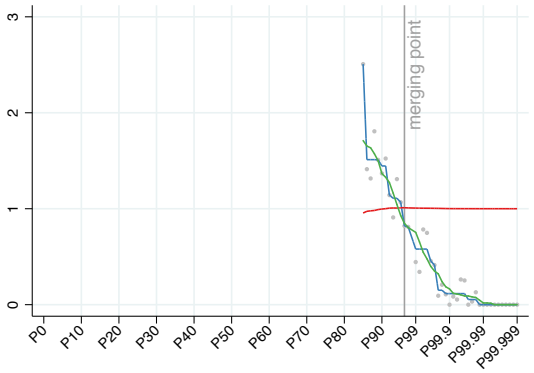
(a) Brazil 2007



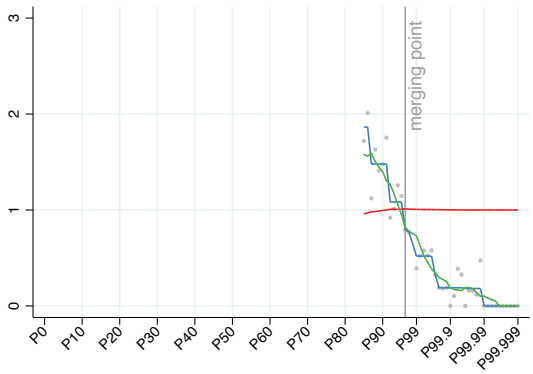
(b) Brazil 2008



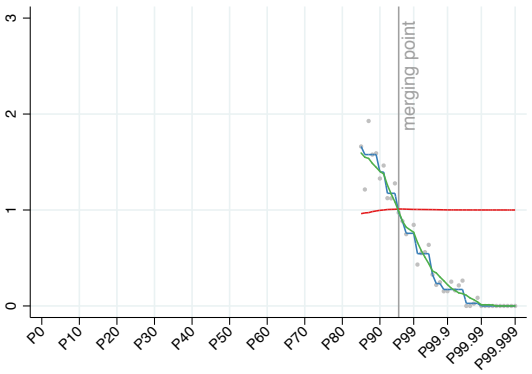
(c) Brazil 2009



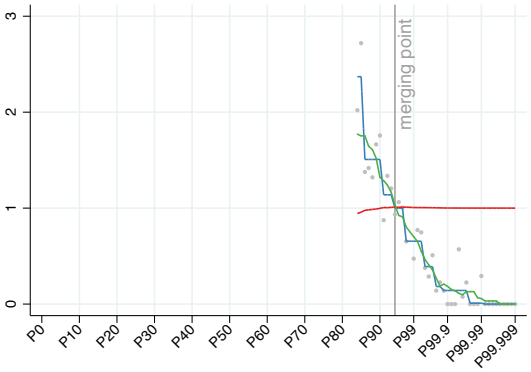
(d) Brazil 2011



(e) Brazil 2012



(f) Brazil 2013



(g) Brazil 2014

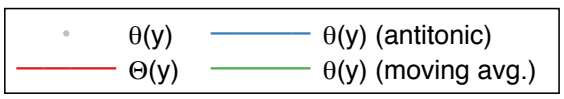
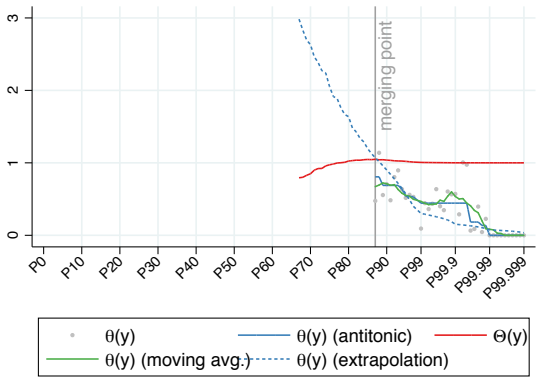
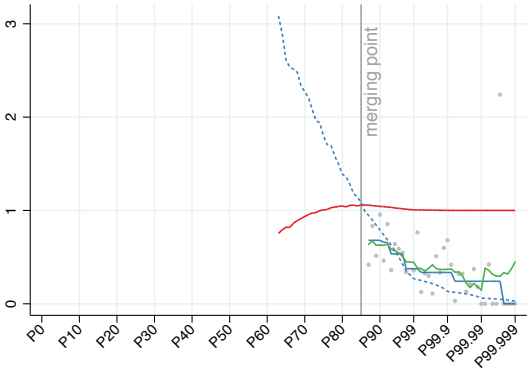


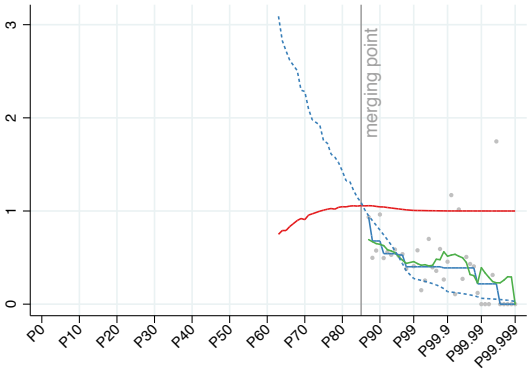
Figure A.5: Merging Points in Chile, 2009-2013



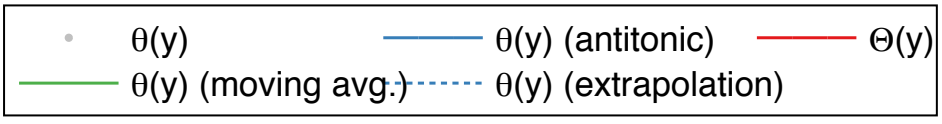
(a) Chile 2009



(b) Chile 2011



(c) Chile 2013



3 Structure of the Corrected Population

Tables A.2-A.6 show the structure of the corrected population for all years in all sampled countries.

Table A.2: Structure of Corrected Population in Brazil, 2007-2015

Year	Population over Merging Point (% total population)		Corrected population		
	Tax data	Survey	Total	Share inside survey support	Share outside survey support
	[2]	[3]	[4] = [2] − [3]	[5]	[6]
2007	1.0%	0.7%	0.33%	98.2%	1.8%
2008	1.0%	0.6%	0.44%	97.2%	2.8%
2009	1.0%	0.5%	0.51%	99.3%	0.7%
2011	2.0%	1.4%	0.57%	95.9%	4.1%
2012	3.0%	2.3%	0.70%	98.3%	1.7%
2013	2.0%	1.4%	0.62%	97.1%	2.9%
2014	2.0%	1.2%	0.76%	98.8%	1.2%
2015	2.0%	1.3%	0.70%	97.2%	2.8%

Notes: Column [2] shows the proportion of the population that is above this merging point in the tax data. Column [3] shows the proportion that is above the merging point in survey data. The difference between the two is the proportion of the survey population that is corrected (Column [4]). As explained in the text, we adjust survey weights below the merging point by the same proportion. The corrected proportion above the merging point can be decomposed into the share of the corrected population that is inside the survey support (up to the survey's maximum income) and the share that is outside the support (observations with income above the survey's maximum).

Table A.3: Structure of Corrected Population in Chile, 2009-2015

Year	Population over Merging Point (% total population)		Corrected population		
	Tax data	Survey	Total	Share inside survey support	Share outside survey support
	[2]	[3]	[4] = [2] − [3]	[5]	[6]
2009	12.0%	7.7%	4.28%	99.7%	0.3%
2011	14.0%	8.9%	5.10%	99.9%	0.1%
2013	14.0%	9.0%	4.98%	99.9%	0.1%
2015	14.0%	9.2%	4.83%	99.99%	0.01%

Notes: Column [2] shows the proportion of the population that is above this merging point in the tax data. Column [3] shows the proportion that is above the merging point in survey data. The difference between the two is the proportion of the survey population that is corrected (Column [4]). As explained in the text, we adjust survey weights below the merging point by the same proportion. The corrected proportion above the merging point can be decomposed into the share of the corrected population that is inside the survey support (up to the survey's maximum income) and the share that is outside the support (observations with income above the survey's maximum).

Table A.4: Structure of Corrected Population in France, 2004-2014

Year	Population over Merging Point (% total population)		Corrected population		
	Tax data	Survey	Total	Share inside survey support	Share outside survey support
	[2]	[3]	[4] = [2] − [3]	[5]	[6]
2004	29.0%	26.8%	2.17%	99.9%	0.1%
2005	25.0%	23.1%	1.95%	98.5%	1.5%
2006	36.0%	32.5%	3.50%	99.5%	0.5%
2007	37.0%	32.0%	4.99%	99.96%	0.04%
2008	0.4%	0.3%	0.11%	97.6%	2.4%
2009	0.1%	0.1%	0.02%	89.8%	10.2%
2010	0.2%	0.1%	0.11%	94.5%	5.5%
2011	0.2%	0.1%	0.06%	94.3%	5.7%
2012	0.2%	0.2%	0.03%	96.5%	3.5%
2013	0.3%	0.3%	0.03%	72.3%	27.7%
2014	0.1%	0.0%	0.05%	99.0%	1.0%

Notes: From 2008, the French survey was supplemented with register data for increased precision in the responses. Column [2] shows the proportion of the population that is above this merging point in the tax data. Column [3] shows the proportion that is above the merging point in survey data. The difference between the two is the proportion of the survey population that is corrected (Column [4]). As explained in the text, we adjust survey weights below the merging point by the same proportion. The corrected proportion above the merging point can be decomposed into the share of the corrected population that is inside the survey support (up to the survey's maximum income) and the share that is outside the support (observations with income above the survey's maximum).

Table A.5: Structure of Corrected Population in Norway, 2004-2014

Year	Population over Merging Point (% total population)		Corrected population		
	Tax data	Survey	Total	Share inside survey support	Share outside survey support
	[2]	[3]	[4] = [2] − [3]	[5]	[6]
2004	24.0%	22.5%	1.49%	99.3%	0.7%
2005	22.0%	19.7%	2.27%	99.8%	0.2%
2006	31.0%	28.8%	2.16%	99.9%	0.1%
2007	39.0%	34.2%	4.75%	99.5%	0.5%
2008	38.0%	33.4%	4.59%	99.95%	0.05%
2009	4.0%	3.5%	0.54%	99.4%	0.6%
2010	8.0%	7.1%	0.88%	99.0%	1.0%
2011	23.0%	21.1%	1.93%	99.0%	1.0%
2012	10.0%	8.9%	1.13%	98.6%	1.4%
2013	22.0%	20.5%	1.49%	99.1%	0.9%
2014	5.0%	4.6%	0.39%	96.0%	4.0%

Notes: Column [2] shows the proportion of the population that is above this merging point in the tax data. Column [3] shows the proportion that is above the merging point in survey data. The difference between the two is the proportion of the survey population that is corrected (Column [4]). As explained in the text, we adjust survey weights below the merging point by the same proportion. The corrected proportion above the merging point can be decomposed into the share of the corrected population that is inside the survey support (up to the survey's maximum income) and the share that is outside the support (observations with income above the survey's maximum).

Table A.6: Structure of Corrected Population in United Kingdom, 2005-2014

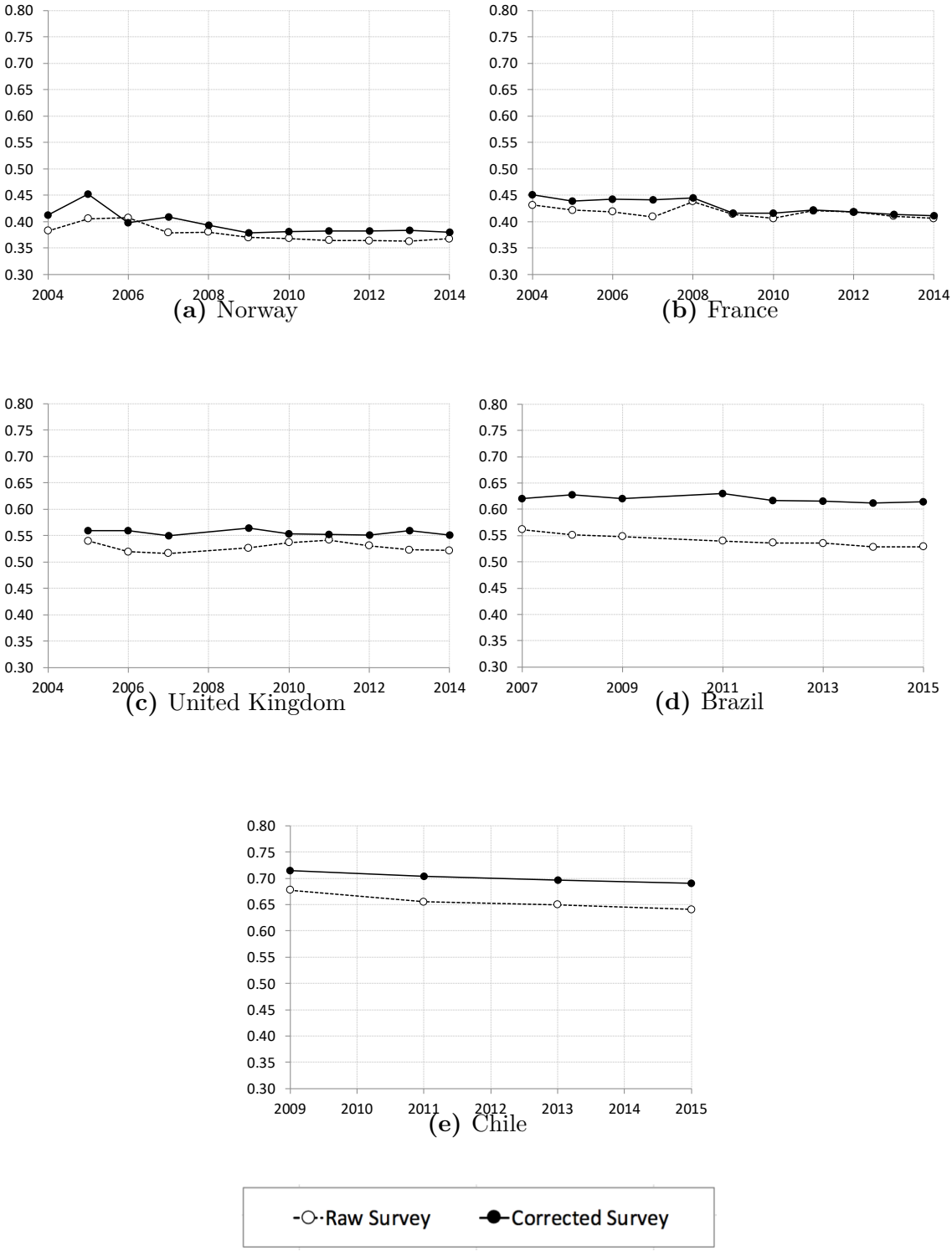
Year	Population over Merging Point (% total population)		Corrected population		
	Tax data	Survey	Total	Share inside survey support	Share outside survey support
	[2]	[3]	[4] = [2] − [3]	[5]	[6]
2005	12.0%	11.7%	0.26%	99.5%	0.5%
2006	8.0%	7.3%	0.72%	96.9%	3.1%
2007	7.0%	6.5%	0.53%	95.5%	4.5%
2009	0.8%	0.5%	0.33%	85.5%	14.5%
2010	0.4%	0.3%	0.14%	84.9%	15.1%
2011	11.0%	10.8%	0.18%	93.0%	7.0%
2012	3.0%	2.6%	0.37%	92.2%	7.8%
2013	4.0%	3.6%	0.45%	86.1%	13.9%
2014	3.0%	2.5%	0.54%	93.6%	6.4%

Notes: Column [2] shows the proportion of the population that is above this merging point in the tax data. Column [3] shows the proportion that is above the merging point in survey data. The difference between the two is the proportion of the survey population that is corrected (Column [4]). As explained in the text, we adjust survey weights below the merging point by the same proportion. The corrected proportion above the merging point can be decomposed into the share of the corrected population that is inside the survey support (up to the survey's maximum income) and the share that is outside the support (observations with income above the survey's maximum).

4 Gini Coefficients

Figure A.6 shows graphs of the Gini coefficients of our 6 country case studies before and after the correction for all available years.

Figure A.6: Gini Coefficients, Before and After Correction in 6 Countries



(a) Norway (b) France

(c) United Kingdom (d) Brazil

(e) Chile

5 The Impact of Misreporting

Our adjustment procedure is based on the interpretation of the whole difference between tax and survey densities as being due solely to nonresponse. However, there are at least some cases in which another co-existing bias is detected. Misreporting in surveys tends to have a negative correlation with income. That is, on average, the poor are more likely to overreport while the rich tend to underreport true income. It is thus fair to ask: what are the consequences of such behavior in our analytical framework?

To define the misreporting bias, let $f_M(y)$ be the distribution of misreported income, $p(y)$ the probability of misreporting for a given level of income and $\bar{p}$ its average. Then we redefine f_Z as the distribution including both the nonresponse and misreporting biases:

$$f_Z(y) = f_Y(y)\theta(y)(1 - p(y)) + f_M(y)\theta(y)\bar{p} \tag{A.1}$$

The left half of the sum refers to those who report income correctly with a given (relative) probability of response ($\theta(y)$). The right-hand side of the sum accounts for those declaring misreported income equal to y, again taking into account the nonresponse bias. In this situation, the over or under-estimation of f_Z with respect to the true distribution (f_Y) can also be formulated as the ratio of the two distributions.

$$\frac{f_Z(y)}{f_Y(y)} = \theta(y)(1 - p(y)) + \frac{f_M(y)}{f_Y(y)}\theta(y)\bar{p} \tag{A.2}$$

If the ratio is higher than 1, the density is overestimated. If it is lower than 1, it is underestimated. Naturally, the shape of such bias depends on the characteristics of each of the variables at play.

The probability of misreporting is likely to be higher in both ends of the distribution, yet we usually do not have explicit information on the actual shape of the misreported distribution. In order to better understand the potential impact that different distributions could cause, it can be useful to analyse a simplified situation where nonresponse does not exist. In that case we would have:

$$\frac{f_Z(y)}{f_Y(y)} = 1 - p(y) + \frac{f_M(y)}{f_Y(y)}\bar{p} \tag{A.3}$$

If misreported income follows the same distribution as true income, that is $f_M(y) = f_Y(y)$, then densities are underestimated where the probability of misreporting is higher than its average ($p(y) > \bar{p}$). Symetrically, densities are overestimated where the same probability is lower than its average ($p(y) < \bar{p}$). Of course, it may seem odd to assume that misreported income is distributed exactly as true income. However, we consider this to be a useful simplification which helps to convey that both the nonresponse and misreporting biases can have a similar impact and that we are unable to tell them apart *ex post*. Indeed, in this case, both biases, either working alone or together, can perfectly

describe a profile as the one in Figure 1.3. If $f_M \neq f_Y$, under some circumstances we can get a similar result. If both densities are of the same type with different parameters (e.g. if both are log-normal with a slightly different mean and standard error) the bias profile would likely have a form similar to Figure 1.3 but with strong or slight perturbations near the mode of each distribution (with unimodal densities). As shown in Section 4.2 our empirical estimate of the θ coefficient, which should be capturing both biases if they exist, describes a rather flat shape through most of the distribution and only falls in the high end of the income distribution where data is sufficient (Norway, France and the United Kingdom). Such a shape implies that, if misreporting has a significant impact on the distribution of survey-income, the differences between f_M and f_Y are not big enough to cause perturbations that are easily distiguishable from noise while observing the θ coefficient. In any case, as far as we know, it is not possible to measure the relative size of both the nonresponse and misreporting biases without access to individually-matched data across datasets.

In a purely distributive perspective, our method is able to correct both the nonresponse and misreporting biases together. Symetrically, we can also find an algorithm that theoretically reproduces the same adjusted distribution *via* modifying individual income. Such an algorithm, would correct for both biases too. Nonetheless, despite having virtually the same distributive results, the reweighting algorithm should be prefered when we aim to use other variables in the survey, because it preserves the internal consistency of each observation. On the contrary, replacing incomes, at least as it has been implemented to date, assumes implicitly a deterministic form of misreporting. That is, everybody underreports increasingly from a given level of income. A correction method that is based on this conception of misreporting most likely modifies the income of those who report correctly, thus, worsening the representativeness of each observation in terms of covariates.

Appendix B

Building DINA for Brazil: Data Appendix

Abstract: This Data Appendix presents a more detailed exposition of the sources and estimation procedure employed in chapters 2 and 3 to build Distributional National Accounts (DINA) for Brazil.

1 The Personal Income Tax in Brazil

1.1 Brief History

The modern federal personal income tax was created in 1922 in Brazil. Nóbrega (2014) offers the definitive account of its evolution over time.

In a nutshell, from its inception until the late 1980s the income tax followed a schedular system of taxation. First, incomes were classified as "gross taxable" and "non-taxable" (that is, either tax-exempt or "taxed exclusively at the source"). Gross taxable incomes were sorted into different income schedules. Then, taxpayers could subtract both schedular deductions and personal allowances (*abatimentos*) in order to obtain their net taxable income, which was subject to progressive tax rates. Until 1964 there were also flat schedular taxes. This system was overhauled in the late 1980s, as the income schedules were abolished and forms and regulations were simplified. Yet, the tax assessment did not change radically. Incomes are still classified as "gross taxable" and "non-taxable"; then, itemized deductions are subtracted from the former to calculate net taxable incomes, which are subject to progressive tax rates.The most consequential changes over time happened in four key definitions: the tax unit, the scope of non-taxable incomes, tax exemption thresholds and marginal tax rates.

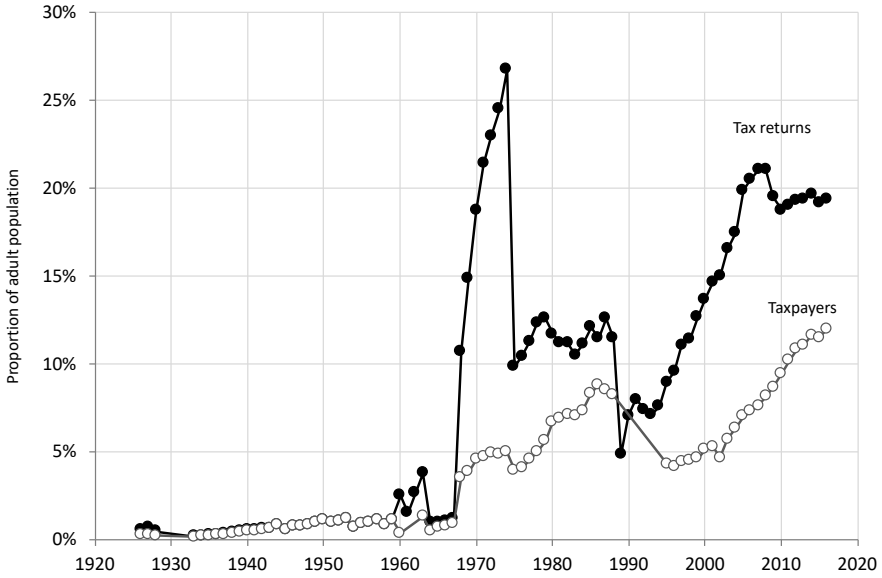
The tax unit was always defined as either couples or individuals. Furthermore, couples almost always could choose whether to file jointly or separately, depending on matrimonial regime and sources of income. In the very early decades the rules were more restrictive and dual-earner families were rare, so most couples likely filed a single tax return. This changed dramatically over time, as women entered the labor force and the tax legislation was updated accordingly. The last restrictions were relaxed in 1994, so couples are now completely free to choose whether to file jointly or not. In practice, couples with two high earners are better off filing separately but otherwise they can lower their tax burden by filing jointly.

Consequently, the definition of the tax unit in Brazil was never as clear-cut as in most countries. In this context it makes sense to estimate an inequality series both for individuals (assuming pure individualistic filing) and equal-split adults (assuming that the income of couples is equally-split).

Capital gains, gifts, inheritances and some government transfers were almost always categorized as non-taxable, but other than that there were many changes both in the scope and the composition of non-taxable incomes. Between the 1920s and the early 1960s, several occupational associations lobbied for, and won, tax exemptions for their earnings, such as judges, teachers, and journalists. Such exemptions were all revoked in 1964. By then, the military dictatorship was focused on granting new exemptions or preferential treatment to business and capital incomes in an attempt to foster savings and investment. Profits and dividends, bonds and other types of investment incomes became, under certain

conditions, either fully exempt or taxed exclusively at the source at lower rates. "Monetary correction" adjustments – that is, the indexation mechanism which protected investments from rising inflation – were also tax free.

Figure B.1: Share of tax returns and taxpayers in Brazil: 1926–2016



Source: Souza (2016, p. 165), updated to 1926 and 2016. Notes: "adult population" comprises individuals aged 20 or above. Taxpayers are tax filers whose net taxable income exceeded the exemption threshold. Tax returns can be filed separately or jointly in the case of married couples. Prior to 1945, when only regional data is available, the number of tax returns for Brazil is estimated using the average ratio of the number of tax returns in the Federal district to the number of tax returns in Brazil between 1945 and 1950. The large fluctuations in tax returns over the late 1960s/early 1970s was due to reforms that changed the income thresholds that obliged individuals to file a tax return.

The macroeconomic stabilisation plan in 1994 marked the end of the "monetary correction", but the scope of non-taxable incomes kept growing. For instance, after 1989 the so-called "13^{th} salary", which had hitherto been fully taxable, became taxed exclusively at the source and separately from all other incomes, at a lower tax rate. More importantly, from 1996, distributed business profits and dividends became exempt from the personal income tax. Previously, they were either fully taxable (until 1974) or taxpayers could choose whether to report them as gross taxable incomes or incomes taxed exclusively at the source. Unfortunately, non-taxable incomes were not even reported on tax forms until the early 1970s. Nowadays, they comprise about 40% of the total income reported on tax returns. This percentage was 10 p.p. lower in the mid-1990s (Souza, 2016, p. 177-178).

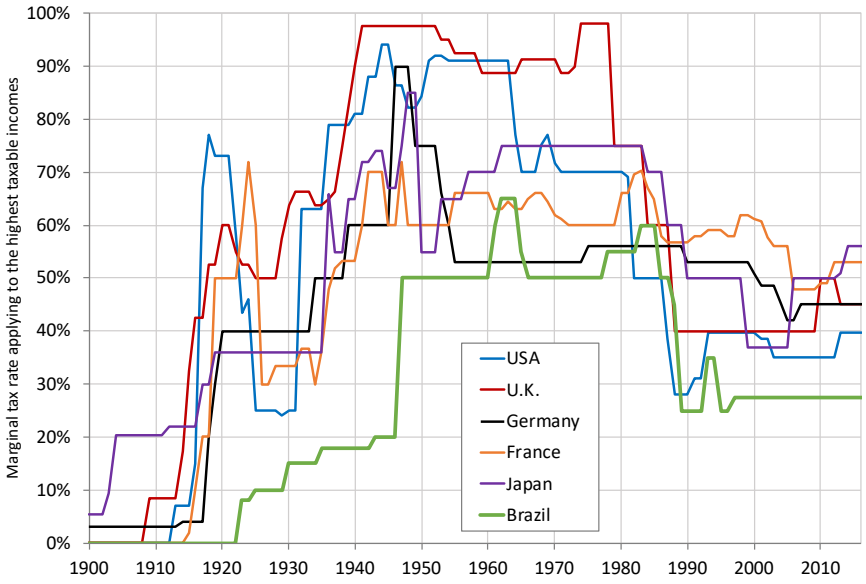
Regarding tax exemption thresholds, as in many other countries, the relative number of income tax filers and taxpayers has been historically very low in Brazil. Figure B.1 shows that until the 1960s, barely 1% of the adult population filed tax returns. This percentage soared briefly as the military dictatorship lowered exemption thresholds, and then hovered

between 10% to 15% during the 1970s and the 1980s. The income tax reforms enacted as the country transitioned back to democracy brought these figures down to 5–10%. Over the past couple of decades they have been on the rise again, reaching about 20% in recent years, mostly due to "bracket creep" – that is, exemption thresholds rising slower than inflation – and the strong period of economic growth in the 2000s.

The share of taxpayers has also risen notably. Early on, there was little difference between assessed tax units and taxable tax units. Since the 1960s, the former outgrew the latter considerably, given the implementation of new policies regarding the assessment of personal incomes. The gap between the two closed in the late 1980s, but again widened after the stabilization of inflation in the mid-1990s, with the numbers of taxable tax units increasing at a slower pace than the numbers of assessed tax units.

Finally, marginal tax rates also varied both in terms of quantity and progressivity. Figure B.2 illustrates the latter point by contrasting the top marginal rates in Brazil and in the United States. Broadly speaking, Brazil followed international trends, with very high top rates in the post-war decades and significant cuts in the 1980s.

Figure B.2: Top marginal tax rates in Brazil and Selected Countries: 1900–2016



Notes: Rates for Brazil are from the *Secretaria da Receita Federal*. Other rates are from OECD Tax Database.

Nonetheless, it is worth repeating that the Brazilian income tax was never as redistributive as Figure B.2 might suggest: on the one hand, as mentioned, the scope of non-taxable incomes rose over time; on the other hand, the spikes in top rates – most notably in 1947 – only impinged on a minuscule number of very, very rich taxpayers. Thus, income shifting and tax evasion are not likely to be hugely problematic.

1.2 List of sources

The publicly available income tax tabulations date back to the 1920s and span almost one century. Generally, they are similar to the statistics published in other countries, listing income totals and the number of tax returns by fine-grained income brackets.

They are not, however, fully consistent over time. Indeed, the Brazilian Tax Agency (*Secretaria da Receita Federal*) and its predecessors published such tabulations intermittently and in a somewhat scattershot fashion, which required a range of procedures to clean the data and harmonize definitions, as discussed below.

For now, it is worth introducing the most unusual feature of the data, namely, that there are no tabulations of total incomes by total income brackets for most of the time. In fact, only from 2006 onwards we have such data.

For the earlier years, the income brackets are ranked mostly by net taxable and/or gross taxable incomes. For the 1980s we also have tabulations ranked by non-taxable incomes. As for the income totals reported, prior to 1974 they are also limited to either net taxable or, at best, gross taxable incomes. As mentioned, the tax forms did not even require taxpayers to disclose their non-taxable incomes until 1970. From 1974 onwards most tabulations also report non-taxable incomes, thus allowing us to calculate total incomes.

Most procedures discussed below were devised to deal with these shortcomings, relying on available data to extrapolate partial incomes (e.g., net taxable incomes or gross taxable incomes) to the relevant income concept, that is, total income. In most cases, the adjustments are rather uncontroversial and could be validated as there are years with multiple tabulations available.

Table B.1 lists all sources with the income tax tabulations we used in this paper.

Table B.1: List of sources for income tax tabulations, 1927–2016

Year	Sources	Coverage[a]	Brackets[b]	Reported incomes[b]
1927	Souza Reis (1930)	DF	Net	Net, Gross
1928	Souza Reis (1930)	DF	Net	Net, Gross
1929	N/A	N/A	N/A	N/A
1930	N/A	N/A	N/A	N/A
1931	N/A	N/A	N/A	N/A
1932	N/A	N/A	N/A	N/A
1933	AEB 1941–1945 (IBGE, 2017)	DF	Net	Net
1934	AEB 1946 (IBGE, 2017)	DF	Net	Net
1935	AEB 1946 (IBGE, 2017)	DF	Net	Net
1936	AEB 1946 (IBGE, 2017)	DF	Net	Net
1937	AEB 1946 (IBGE, 2017)	DF	Net	Net
1938	AEB 1946 (IBGE, 2017)	DF	Net	Net
1939	AEB 1946 (IBGE, 2017)	DF	Net	Net
1940	AEB 1946 (IBGE, 2017)	DF	Net	Net
1941	AEB 1946 (IBGE, 2017)	DF	Net	Net
1942	AEB 1946 (IBGE, 2017)	DF	Net	Net
1943	AEB 1946 (IBGE, 2017)	DF+3	Net	Net
1944	Mortara (1949b, p. 14)	Brazil	Net	Net
1945	AEB 1948 (IBGE, 2017)	Brazil	Net	Net
1946	AEB 1949 (IBGE, 2017)	Brazil	Net	Net

Continues on the next page

Table B.1. *Continued from previous page*

Years	Sources	Coverage[a]	Brackets[b]	Reported Incomes[b]
1947	AEB 1949 (IBGE, 2017)	Brazil	Net	Net
1948	AEB 1950 (IBGE, 2017)	Brazil	Net	Net
1949	AEB 1951 (IBGE, 2017)	Brazil	Net	Net
1950	AEB 1952 (IBGE, 2017)	Brazil	Net	Net
1951	AEB 1953 (IBGE, 2017)	Brazil	Net	Net
1952	AEB 1954 (IBGE, 2017)	Brazil	Net	Net
1953	AEB 1955 (IBGE, 2017)	Brazil	Net	Net
1954	AEB 1956 (IBGE, 2017)	Brazil	Net	Net
1955	AEB 1957 (IBGE, 2017)	Brazil	Net	Net
1956	AEB 1958 (IBGE, 2017)	Brazil	Net	Net
1957	AEB 1959 (IBGE, 2017)	Brazil	Net	Net
1958	AEB 1960 (IBGE, 2017)	Brazil	Net	Net
1959	AEB 1961 (IBGE, 2017)	Brazil	Net	Net
1960	AEB 1962 (IBGE, 2017)	Brazil	Net	Net
1961	N/A	N/A	N/A	N/A
1962	N/A	N/A	N/A	N/A
1963	Departamento do Imposto de Renda (1965)	Brazil	Net	Net
1964	Divisão do Imposto de Renda (1966)	Brazil	Net	Net, Gross
1965	Departamento do Imposto de Renda (1968a)	Brazil	Net	Net, Gross
1966	Departamento do Imposto de Renda (1968b)	SP+GB	Net	Net, Gross
1967	Centro de Informações Econômico–Fiscais (1968)	Brazil	Net	Net, Gross
1968	AEF 1970 (Secretaria da Receita Federal, 1980)	Brazil	Net	Net, Gross

Continues on the next page

Table B.1. *Continued from previous page*

Years	Sources	Coverage[a]	Brackets[b]	Reported Incomes[b]
1969	AEF 1971 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross
1970	AEF 1972 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross
1971	AEF 1973 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross
1972	AEF 1974 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross
1973	AEF 1975 (Secretaria da Receita Federal, 1980)	Brazil	Gross	Net, Gross
1974	AEF 1976 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross, Non-Tax., Total
1975	AEF 1977 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross, Non-Tax., Total
1976	AEF 1978 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross, Non-Tax., Total
1977	AEF 1979 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross, Non-Tax., Total
1978	AEF 1980 (Secretaria da Receita Federal, 1980)	Brazil	Net, Gross	Net, Gross, Non-Tax., Total
1979	IRPF 1980 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1980	IRPF 1981 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1981	IRPF 1982 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1982	IRPF 1983 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1983	IRPF 1984 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1984	IRPF 1985 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1985	IRPF 1986 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1986	IRPF 1987 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1987	IRPF 1988 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1988	IRPF 1989 (Secretaria da Receita Federal, 1989)	Brazil	Net, Gross, Non-Tax.	Net, Gross, Non-Tax., Total
1989	N/A	N/A	N/A	N/A
1990	N/A	N/A	N/A	N/A

Continues on the next page

Table B.1. *Continued from previous page*

Years	Sources	Coverage[a]	Brackets[b]	Reported Incomes[b]
1991	N/A	N/A	N/A	N/A
1992	N/A	N/A	N/A	N/A
1993	N/A	N/A	N/A	N/A
1994	N/A	N/A	N/A	N/A
1995	N/A	N/A	N/A	N/A
1996	Secretaria da Receita Federal (2001)	Brazil	Gross	Gross
1997	Secretaria da Receita Federal (2001)	Brazil	Gross	Gross
1998	Secretaria da Receita Federal (2000)	Brazil	Gross	Net, Gross
1999	N/A	N/A	N/A	N/A
2000	Torres (2003)	Brazil	Gross	Gross
2001	N/A	N/A	N/A	N/A
2002	Secretaria da Receita Federal (2004)	Brazil	Gross	Gross
2003	N/A	N/A	N/A	N/A
2004	N/A	N/A	N/A	N/A
2005	N/A	N/A	N/A	N/A
2006	Medeiros et al. (2015a)	Brazil	Total	Gross, Non-Tax., Total
2007	GNIRPF 2007 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2008	GNIRPF 2008 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2009	GNIRPF 2009 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2010	GNIRPF 2010 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2011	GNIRPF 2011 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2012	GNIRPF 2012 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total

Continues on the next page

Table B.1. *Continued from previous page*

Years	Sources	Coverage[a]	Brackets[b]	Reported Incomes[b]
2013	GNIRPF 2013 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2014	GNIRPF 2014 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2015	GNIRPF 2015 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total
2016	GNIRPF 2016 (Secretaria da Receita Federal, 2018)	Brazil	Net, Gross, Non-Tax., Total	Net, Gross, Non-Tax., Total

Note: "N/A" means "Not Available".

[a] "DF" stands for the former Federal District (*Distrito Federal*), now the municipality of Rio de Janeiro. "DF+3" means the former Federal District plus the states of São Paulo, Minas Gerais and Rio Grande do Sul. "SP+GB" means São Paulo and the former state of Guanabara (the municipality of Rio de Janeiro in 1960–1975).

[b] "Net" means net taxable income; "Gross" means gross taxable income; "Non-Tax." means non-taxable income (either exempt or taxed exclusively at the source); "Total" means total income. Most sources between 1947 and 1988 also have information on *renda bruta*, i.e., gross taxable incomes minus schedular deductions.

2 The Brazilian Household Survey

Brazil's flagship household survey, the Pesquisa Nacional por Amostra de Domicílios (PNAD), is a large, multi-purpose, nationally representative survey run by the Instituto Brasileiro de Geografia e Estatística (IBGE, the Brazilian National Statistics Bureau).

The PNAD started in 1967 as a small-scale quarterly survey with limited coverage. After two interruptions and numerous refinements, it was relaunched in 1976 and ran annually until 2015, except in Census years (1980, 1991, 2000 and 2010). It also did not run in 1994 due to budget cuts. Martine et al. (1988), Médici (1988) and Martine (2005) provide detailed historical accounts of the origins and evolution of the PNAD over time.

The PNAD was gradually improved over the years, but there were no marked breaks in terms of variables and definitions. Almost all years include comparable information on labor market outcomes, incomes, education and so on (see, for instance, Soares and Lima (2002), on the educational variables and Médici (1988) and Rocha (2003) on incomes).

The PNAD was discontinued in 2015 as IBGE transitioned to PNAD Contínua (PNADC), a quarterly survey with a rotating panel similar to the United States's Current Population Survey (CPS). Both the sampling design and the questionnaire were overhauled, so the PNADC results are not directly comparable to the historical PNAD series for most variables.

We use the individual-level microfiles provided by IBGE for the PNAD between 1976 and 2015 to extract personal incomes. The number of income variables rose over time from 8 in 1976 to 14 from 1992 onwards. Generally speaking, this information has good quality and a few unsurprising shortcomings (Médici, 1988; Rocha, 2003; Souza, 2015). As usual, the PNAD underestimates top incomes and business and investment incomes (Hoffmann, 1988; Souza, 2015; Morgan, 2017).

Sample sizes were always very large. The PNAD used a probability selected sample of about 115,000 households and 388,000 individuals per year. There was some variation over time as the samples were slightly smaller in the late 1980s. Overall, the PNADs range from 65,000 to 154,000 households and from 290,000 to 534,000 persons per year.

The data are nationally representative from 2004 onwards. Previous years do not include rural areas of six northern states (Rondônia, Acre, Amazonas, Roraima, Pará, and Amapá). Moreover, the data for 1976-1979 also excludes rural areas in center-western states (Mato Grosso, Mato Grosso do Sul, Goiás and Tocantins, which was still part of Goiás). According to the 1980 and 1991 Censuses and the recent PNADs, this amounts to excluding 3-4% of the population in the 1970s and about 2% between 1981 and 1990.

We make no adjustments to account for this missing population, given that the distributional statistics hardly change when we compare the full sample to the restricted sample in 2004 and perform an extrapolation for earlier years, weighted by the excluded population share.

3 National Accounts in Brazil

Official national accounting harkens back to the late 1940s. Nunes (1988) and Hallak Neto and Forte (2016) recount the evolution of the Brazilian System of National Accounts in detail. Their research can be summarized as follows.

Briefly, several scholars had already tried to calculate historical series for GDP and other macroeconomic aggregates, but the first official figures were estimated by Fundação Getúlio Vargas (FGV-RJ) for the year 1947. At that point the estimates were limited to aggregates such as net national income, and then gradually expanded. FGV-RJ constantly updated its methodology and tried to adapt to international guidelines, mostly the SNA-53.

Still, the dearth of primary sources meant all figures could be consistently estimated solely for selected years and the gaps were filled by extrapolation from more general indicators on the level of production and prices. Also, constant methodological refinements entailed some breaks in the comparability of the series.

The most serious shortcoming, however, was that FGV-RJ's approach was focused on institutional sectors and did not present input-output tables, contrary to SNA-68 guidelines. Hence, IBGE started its national accounting project to calculate such tables, publishing the first results in the late 1970s.

Consequently, Brazil was left with two competing and not fully coherent systems of national accounts. For years, even GDP estimates differed between the FGV and IBGE results. This conundrum came to an end in 1986, when IBGE became fully responsible for Brazil's national accounts. This led to estimates for 1980-1993 (base-year 1980) which followed the SNA-68 more closely and paved the way for a new series for 1985-2003 (base-year 1985) based on the SNA-93.

National account statistics were overhauled again in 2007 (base-year 2000) and 2015 (base-year 2010) in order to take into account the latest international guidelines and also to make better use of the data collected by other surveys administered by IBGE. The new series no longer relied on production and price indexes to extrapolate the data originally calculated for a single base-year. Such indicators were used only to retropolate results for earlier years. Thus, the SNA-93 series for base-year 2000 was stretched as far back as 1995 and the SNA-08 series for base-year 2010 was retropolated back to 2000.

4 Data Preparation

The preparation of the data for building Distributional National Accounts (DINA) needs to account for the different degrees of income coverage in each source. The income concept we wish to finally adhere to is *pre-tax post-replacement national income.* That is, we are interested in income before the payment of income taxes, but including "replacement" income transfers such as pensions and unemployment insurance, for which workers have to save for out of their current income. This income must therefore be net of contributions

made to save for these social insurance transfers. While this income concept can, by definition, be constructed from national accounts, it cannot be directly composed from our two main sources of distributional data, the survey micro-files and personal income tax tabulations.

What we call “survey income", does not account for imputed rent or social contributions paid out of wages/salaries. It is also notably thin on property incomes as compared to the tax data (see chapter 2). The latter, while only covering a smaller fraction of the population, better captures most of the nations' capital income when tabulations of total (taxable and tax-exempt) income exist. Yet it still does not report imputed rent, given its non-assessed status for fiscal purposes. The tabulated tax data is also gross of social contributions. Furthermore, both sources of distributional data do not include incomes accrued to households, or over which households are the beneficial owners, such as investment income on pension and insurance funds accruing to households and the undistributed profits of corporations. For these reasons we must proceed gradually towards our goal of distributing national income. The intermediate steps will prove useful to compare how income is distributed in these different dimensions. In the first instance, we must reconcile the Brazilian surveys with the tax data so as to get a more complete picture of the distribution of pre-tax post-replacement “fiscal income" received by households. And for this, certain modifications to the raw tax and survey data are required. We explain these in turn.

4.1 Tax Tabulations

As shown in Table B.1, throughout the period from the mid-1920s to the present, distinct types of income tabulations are publicly available. They cover separate, and sometimes overlapping, income concepts. Thus, some adjustments and imputations to the raw data are required.

Partial geographic coverage in 1926/1927–1943 and 1966

Our income tax tabulations do not have national coverage from 1927 to 1943 and in 1966. Until 1942, the tabulations report only tax returns for the former Federal District, which is now the municipality of Rio de Janeiro. The figures for 1943 refer to the states of São Paulo, Minas Gerais, and Rio Grande do Sul. The data for 1966 covers only the states of São Paulo and Guanabara (again, now known as the municipality of Rio de Janeiro).

Albeit unfortunate, such limitations are not too severe, considering these states raised the lion's share of the income tax revenue back then. Between 1925 and 1942, the Federal District accounted for a higher share of the personal and corporate income tax than any other state (about 38%). In 1943 the four states (including the Federal District) covered by the tabulations raised more than 80% of the income tax revenue. Finally, São Paulo and Guanabara accounted for 72% of the income tax revenue in 1966.

The adjustment to scale-up the partial coverage to national figures was straightforward. For 1927–1942, we multiplied the raw data by $1/revenue_{DF}$, that is the inverse of the share of the Federal District in the total income tax revenue in each year. The adjustment for 1943 was defined analogously as $1/revenue_{DF,SP,MG,RS}$. For 1966, the multiplier was slightly different: $1/taxpayers_{SP,GB}$, that is, the inverse of the share of taxpayers living in São Paulo and Guanabara.

We validated this approach by comparing scaled-up regional estimates with national figures for 1945, 1946 and 1947. The results were very close: the scaled-up tabulations typically differed from observed results by less than 5%. Moreover, none of the fiscal top incomes series show spikes or discontinuities that could be traced back to the scaling-up adjustment.

Number of tax returns per income brackets in 1963 and 1964

The income tax tabulations for 1963 and 1964 do not report the number of tax returns per income brackets. In turn, the figures reported by Kingston and Kingston (1972) for 1964 are not fully reliable, as two intermediate income brackets have average incomes outside the valid range.

Therefore, we proceeded as follows: for both years, average incomes were imputed as the midpoint of each income bracket, as roughly observed in neighboring years with more reliable data. For the top brackets, we assumed an inverted Pareto coefficient (β) of 1.8 and imputed the average income and number of tax returns accordingly.

The imputations are very close to the number of tax returns cited by Kingston and Kingston (1972), except, of course, for the two aforementioned intermediate brackets. Once again, this adjustment does not introduce any unwarranted breaks in the top incomes. Considering the top brackets comprised only a minuscule number of tax returns in the 1960s, changing the inverted Pareto coefficient does not alter our results.

Imputation of gross taxable incomes, 1926/1927-1963

The publicly available income tax tabulations do not report gross taxable incomes before 1964; instead, they mostly present totals per bracket for net taxable incomes, that is, gross taxable incomes minus schedular deductions and personal allowances. Following Souza (2016, p. 192–193), the procedures to extrapolate to gross taxable incomes differed according to the available data:

- 1926/1927–1928: we assumed all brackets had the same gross/net income ratio as observed for the aggregate total (1.49).
- 1933–1946: first, we imputed *renda bruta* by multiplying each bracket's net taxable income by 1.25 (which implies personal allowances were 20% of the *renda bruta*); second, we imputed gross taxable incomes by multiplying each bracket's imputed *renda bruta* by 1.13 (assuming schedular deductions were 12% of gross taxable

income).

- 1947–1960: first, we assumed the composition of net schedular incomes was the same across all brackets; second, we applied multipliers of 1.21 (schedules A and B), 1.14 (schedule C), 1.64 (schedule D), 1.33 (schedule E), 1 (schedules F and G) and 1.17 (schedule H) to convert net schedular incomes into gross schedular incomes, which were summed to yield gross taxable incomes.

- 1963: each bracket's observed *renda bruta* was multiplied by 1.19.

All multipliers were based on the closest neighboring years with the relevant data and fine-tuned to take into account major changes in the tax laws pertaining to allowances and deductions. After these procedures, all income thresholds were adjusted by the resulting gross/net income ratio.

Fixed capital consumption, 1926/1927-2016

In order to make incomes from the tax data fully consistent with the concept of "fiscal income" we wish to capture at this stage, we need to subtract fixed capital consumption of unincorporated business from gross taxable incomes, that is, business expenses of independent workers required to keep accountancy books (e.g. doctors, dentists, psychologists, lawyers, independent commercial agents, etc.), as these expenses are incurred to generate their income. Such expenses are not identifiable in the household survey, but we know from the tax statistics that these generally affect higher incomes more, which the fiscal data does better to capture.

From 1926/1927 to 1988, we subtracted 50% of deductions reported on schedules D (income from non-salaried, professional occupations) and E (incomes from real estate activities, including rentals), i.e., we took a conservative stance as deductions were very broadly defined back then. For years with no breakdown by income schedules we set fixed capital consumption as 6% of gross taxable incomes for each bracket, as was roughly the case in years with appropriate data.

After the income tax reform of the late 1980s, fixed capital consumption became the so-called "livro-caixa" deduction, which was subtracted from total assessed income whenever this information was available; otherwise we imputed them using bracket-specific percentages, varying from 0.05% to 4% of gross taxable incomes, estimated from years with complete information.

Disentangling 1926 from 1927

Souza Reis (1930, p. 27) warns the tabulations he presents for 1927 conflate incomes earned in 1926 and 1927. Unfortunately, he makes no attempt to separate incomes from both years. Nevertheless, elsewhere in the book he mentions that about 25% of the income tax revenue in the Federal District in 1928 was collected on incomes earned in previous

years (Souza Reis, 1930, p. 23). Finally, he also shows total revenue from the income tax increased 16.63% from 1926 to 1927 (Souza Reis, 1930, p. 36).

Following Souza (2016; 2018), we make use of these figures to adjust the numbers for 1927 and to estimate an income tax tabulation for 1926. First, we assume the figures for 1927 are also inflated by 25%, so we multiply them by 0.75 to obtain the adjusted tabulation for that year. Second, we assume the distribution of income in 1926 was the same as in 1927 and that all brackets had the same nominal income growth rate of 16.63%. Consequently, we obtain the income tax tabulation for 1926 by multiplying the adjusted 1927 figures by 0.8574 (1/1.1663).

4.2 Estimating the Share of Individual Tax Declarations

In Brazil, the tax unit is the adult individual or married couple, in cases when spouses opt to declare jointly. This has been the case since the birth of the income tax. This decision depends on the income differences between individuals in a couple. A jointly filed declaration takes the combined total income of the couple for the application of the tax schedule. Where a spouse has little income relative to his/her partner, there are more incentives to file jointly if the fixed allowance for dependents (including spouses with or without income) that is deductible from gross income is greater than the additional tax burden brought about by a joint declaration. But if the spouse has higher income then incentives increase for her to file separately, as her income would be subject to the different marginal tax rates (including the first exempt threshold) as opposed to being all subject to the highest rate if she filed jointly with her partner. Given that we aim to distribute individual income, we are forced to make certain assumptions in order to estimate individual declarations from those that can be returned jointly or separately by couples. This will allow us to equally-split income on the estimated portion of declarations that are jointly filed.

Since joint filing by couples is voluntary, this brings forth the complication that not all single declarations are made by persons who are actually single in their civil status. Some couples may choose to file separate declarations for the fiscal reasons outlined in the previous paragraph. For the years 1968-1988, this is not an issue as the tabulations present bracket information on the civil status of filers, which allows a precise calculation of the share of actual single persons per bracket. However, for the 2007-2016 this breakdown by civil status is no longer presented so we rely on a different estimation strategy, as we explain below. For the latter period, our estimation assumes that individuals in couples filing separately are either single or married to other individuals whose income falls in the same bracket, which may not be true. This means that we may over-state inequality compared to the perfect equal-split case (where the total income of actually existing couples is divided by 2), and may under-state inequality as compared to the pure individualized case (where each spouse is assigned his or her own income). Similarly, the estimation of

a pure individualized series is complicated by the fact that we do not know the income composition of jointly-filing couples. These limitations can only be remedied by having access to the Brazilian income tax declaration micro-files.

For the different income concepts across different years that we have tabulated income data, we estimate the share of single declarations per bracket as follows:

Total income For the 2007-2016 period, we use the total value of the deduction for dependents per bracket in the tabulations (Table 9 of the published statistics) and its fixed value per dependent defined in the tax law to calculate the number of dependents per bracket. This number includes spouses, children and other relatives. In order to calculate the number of spouses appearing on a joint declaration, we use the share of spouses in total dependents per bracket of household heads' income from the PNAD surveys. This share varies from about 25% for the lowest bracket to 40% for the highest bracket. Given the condition that persons filed as dependents (with or without income) on a declaration cannot file a separate tax return, the resulting estimation gives us the share of single declarations per bracket, such that we can calculate the equal-split adult income series. The share of single declarations generally falls with income. Overall, joint declarations make up about 20-25% of all the filed declarations. For 2006 we apply the same bracket share of singles estimated for the 2007 tabulation, average across values for top brackets to fit the reduced brackets of the 2006 tabulation.

Gross taxable income For 2007-2016, we follow the same procedure as the one applied to total incomes but apply it to the gross taxable income tabulation (Table 7 of the published statistics). For the years 1969-1988, we use the information on the number of declarations made by civil status (single, married, widowed, other) and gender to directly compute the share declarations made by single men and women in the total number of declarations. During these years single declarations (of single adults or married spouses) made up 30% of all tax returns, with the remaining 70% being the share of jointly filed declarations.[1] For years without such information, we calculate the bracket shares as the averages of the bracket shares of the closest years. For 1998 we calculate the share of singles per bracket as the average of the shares for 1988 and 2007. For 2002 we apply the average of the 1998 and 2007 bracket shares of single declarations. The values for 2000 are the bracket averages of 1998 and 2002 (adjusted for the greater number of brackets), while the values for 1996 and 1997 are assumed to be equal to the values observed for 1998.

[1] Unfortunately, we cannot known the income composition of couples in jointly-filed declarations. We can only distinguish the number of married women filing a tax return from non-married women. Unsurprisingly, the number of joint-filers headed by men has always been greater for the years in which we have a breakdown by civil status and gender (1969-1988). However the number of women filing a tax return has grown five fold over the period, from 4% in 1969 to 21% in 1988. Thus women have substantially increased their share in joint declarations, at least up until the late 1980s.

Net taxable income For 2007-2016: we follow the same procedure as the one applied to total and gross taxable incomes but apply it to the net taxable income tabulation (Tabela 6 of the DIRPF Large Numbers). For the years 1968-1977, we use the information on the number of declarations made by civil status (single, married, widowed, other) and gender to directly compute the share declarations made by single men and women in the total number of declarations. For years without such information, we calculate the bracket shares as the averages of the bracket shares of the closest years. For 1926-1967, since no data on the decomposition of declarations by civil status or gender is available for this period, we apply the bracket shares of singles from 1968 to all previous years. For tabulations with more or less brackets than the 1968 tabulation, we adjust the bracket share of singles accordingly (i.e. averaging across brackets for tabulations with less brackets than 1968, and applying the last bracket share of singles to all additional brackets, where the number of brackets is higher than in 1968).

4.3 Survey Microdata

The PNAD collects information on gross monthly incomes by source, recording separately the figures for the major sources (labor market earnings, pensions, and so on) and lumping the rest together in a residual "other incomes" variable. Following Morgan (2017), we harmonize the surveys with our definition of fiscal income by annualizing observed incomes (which includes the imputation of infrequent payments) and disaggregating the "other incomes". Thus, we reclassify PNAD incomes into 10 categories:

1. Employees's incomes: wages from main occupation for formal and informal employees + earnings from other occupations + imputed holiday bonus + imputed 13^{th} salary.
2. Employers's incomes: earnings from main occupation for employers.
3. Self-employment incomes: earnings from main occupation for self-employed individuals with no employees.
4. Pensions: public and private pensions + imputed 13^{th} salary for public pensioners.
5. Welfare transfers: the *Renda Mensal Vitalícia* (RMV, 1976-1995), the *Benefício de Prestação Continuada* (BPC, 1996-2015) and the *Bolsa Família* and its predecessors (BF, 2001-2015) had to be separated from the residual "other incomes" variables. Thus, until 1995 we assumed "other incomes" equal to 50% of the current minimum wage were RMV transfers; after 1996, "other incomes" equal to 100% of the current minimum wage were BPC transfers; and we applied a variation of the "typical values" method to identify BF transfers since 2001 (Soares et al., 2010; Souza, 2013).
6. Financial incomes: for 1977-1979, financial incomes; for all other years, "other incomes" (minus welfare transfers) greater than the current minimum wage.

7. Rental income: observed real estate incomes received by landlords.

8. *Abono de permanência*: unavailable information in the 1970s; observed income from 1981 onwards.

9. *Abono salarial*: non-existent until 1989; imputed as one minimum wage for eligible formal private sector employees from 1990 onwards.

10. Unemployment insurance: non-existent until 1986; imputed from 1987 onwards for respondents who claimed to have received unemployment benefits at some point in the 12 months before the PNAD interview. Benefit levels were imputed as yearly averages.

For the factor income decompositions, we classified (1), (4), (8), (9) and (10) as "labor", (3) as "mixed", and (6) and (7) as "capital". In our benchmark scenario, employers's incomes (2) up to the limit of the income tax exemption were assigned to "labor" whereas excess incomes above the exemption threshold were considered returns to "capital". We also calculated two alternative scenarios using the current minimum wage and the maximum *salário de contribuição* ("contributing salary") to Social Security as thresholds.

Our survey estimates discard all individuals younger than 20 years old and/or with missing income information. Also, our annualization factor did not simply multiply observed incomes by 12 due to the runaway inflation which prevailed during most of the 1976–2015 period. Instead, we took into account both monthly inflation rates and the reference months of the survey (September, usually) to calculate more precise annualization factors. These factors are the product of:

$$\sum_{i=1}^{n=12} \frac{\pi_i}{\pi_{i=PNAD}} \tag{B.1}$$

where π_i is the inflation rate of month i and $\pi_{i=PNAD}$ is the inflation rate of the month during which the PNAD survey was conducted. Our estimates range from 9.30 in 1992 to 12.01 in 1998. As expected, all multipliers after the macroeconomic stabilization in 1995 are very close to 12.

We also impute estimate the distribution of non-fiscal incomes like imputed rents and Social Security contributions per adult, which will be useful for the later steps in the DINA estimation. Social contributions were straightforward: analogously to the imputation of labor benefits, we apply prevailing rates to survey variables such as earnings and employment status, with the aid of a few simplifying assumptions. We use annual historical data from the Secretariat of Social Security of the Ministry of Finance (*Secretaria de Previdência*) on contributory schedules, including the minimum contributing salary (minimum wage), maximum contributing salary and the rates by employment status.[2] For

[2] http://www.previdencia.gov.br/dados-abertos/dados-abertos-previdencia-social/.

imputed rents, we estimate a per adult value for each household using the household survey wave and information on rental values and dwelling characteristics, which we then merge to the individual survey wave. Specifically, we adopt the standard two-step approach in the literature (Ferreira et al., 2003; Morais and Cruz, 2003; Soares, 2017): first, we estimated hedonic rental price regression models for rented dwellings; then we apply the coefficients to owner-occupied households to predict imputed rent values.

5 Estimating Top Income Shares from Tax Data

Our method for building DINA uses distributional information from the surveys, whose microdata is accessible from 1976. To go back further in time we must makes use of information in the tax data along with national accounts. As we explain in section 6.7, our DINA estimates for the top of the distribution are anchored to our estimates of top income shares from the traditional approach. This approach makes use of the adjusted tabulations described in the previous section along with two exogenous parameters - a reference total for the adult population (given the historically low percentage of tax filers in the population) and a reference total for income (given the grouped nature of the data in the tabulations) - as well as the choice of an interpolation method to obtain the desired fractiles. We construct our benchmark series of total income by joining the most complete estimates from different tabulations, whenever there were multiple available for the same years.

5.1 Reference Total for Population

Due to only a small share of the population filing a tax return (see B.1), an exogenous control is needed to define the number of tax filers in each fractile. There are two aspects to this choice. First, one must define a proper unit of analysis, which in turn depends on how the tax law treats married couples and families. Second, one must establish an age cut-off to delimit the adult population.

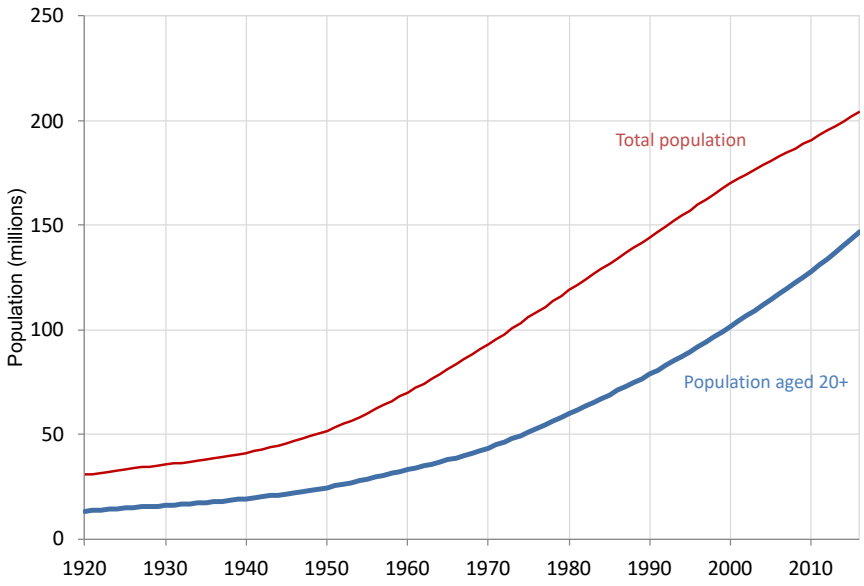
In Brazil, the tax legislation was always somewhat ambiguous regarding the tax unit. Most of the time, married couples were encouraged to file jointly, but all restrictions were relaxed in the early 1990s. In this paper, we define the tax unit strictly as individuals, similar to Medeiros et al. (2015a) and Morgan (2015; 2017) and Souza (2015; 2016). We do so in order to be consistent with the unit of observation in the numerator. As described in Section 4.2, we estimate the shares of single declarations and joint declarations in total declarations to equally split income among spouses in the latter case. This allows us to distribute individual income in a context where still up to one quarter of all declarations have been filed jointly by couples in recent years. In Section 5.4 we also present results when we assume that each declaration corresponds to one income recipient.

We set age cut-off at 20 years old, that is, the population denominator corresponds to

all individuals aged 20 or older. This is in line with most country studies, which normally set the cut-off somewhere between 15 and 21 years (Atkinson et al., 2011). As shown by Atkinson and Piketty (2007), for a given Pareto coefficient α, the income share of the top x% changes by a factor of $(1+c)^{\frac{\alpha-1}{\alpha}}$ if one multiplies the control for total population by (1 + c). Roughly speaking, this implies that the income share of the top 1% in Brazil would be 7% (1.5 p.p.) higher if one lowered the age cut-off from 20 to 15.

The population data come from the decennial census carried out by IBGE, Brazil's central statistics office. For intervening years, estimates were computed via cubic spline interpolation. Figure B.3 plots the total population and the chosen population denominator over time. Brazil faced a very swift demographic transition - not only did the total population grow six-fold between 1925 and 2015 (from 33.2 to 199.8 million) but its age composition also changed dramatically. Adults aged 20 or above were about 45% of the total population in the 1920s and are now 70%.

Figure B.3: Total Population and Reference Total for Adult Population: Brazil 1920–2016



Notes: Authors' calculations based on the 1920, 1940, 1960, 1970, 1980, 1991, 2000 and 2010 Decennial Censuses by IBGE. Figures for inter-census years are computed by cubic spline interpolation.

5.2 Reference Total for Income

In the literature on estimating top income shares, it is common to use an external reference total for income, so that the incomes observed in the tax data can be expressed as a share of an equivalent total. We stress that in our methodology, the reference total for income is used purely for comparative purposes, in relation to the total income we get from combining

incomes in PNAD and IRPF. Indeed, the quality of national accounts deteriorates the further we go back in time, especially in less-developed countries. This warrants a more careful analysis of the income recorded in all available sources to determine the appropriate size distribution of income.

The most common approach to define a control total for income is what Atkinson and Piketty (2007) calls the “top-down" approach. Starting from the national accounts, researchers construct income totals comparable to what is reported on income tax returns, effectively treating the incomes of non-filers as a residual. One possible problem is that detailed national accounts might not be available for the entire period of interest. In this case, the standard solution is to compute the proper income denominator for all possible years and then anchor it to GDP, that is, average it out and use a constant percentage of GDP for all remaining years. Most previous work on Brazil followed this approach. For instance Medeiros et al. (2015a) set the control for total income at roughly 67% of GDP, whereas Morgan (2015) defined it as 60% of GDP. In this paper, we refine this approach. As noted by Atkinson and Piketty (2007), it is very unlikely that a constant percentage is the optimal choice, considering the expansion of the welfare state and related changes over the 20th century. Thus, following Souza (2015; 2016), we start from the most recent national accounts data, which are the most detailed, to estimate the reference total for pre-tax fiscal income as follows (SNA codes are in brackets):

Total pre-tax fiscal income (2000-2015)
= Salaries (D11, S14)
+ Gross operating surplus (B2, S14)
+ Gross mixed income (B3, S14)
+ Gross interests (D41 resources, S14)
+ Dividends and other withdrawals (D42, S14)
+ Investment income attributed to investment fund shareholders (D443, S14)
+ Social security benefits in cash (D621 + D622, S14)
+ Gross current transfers (D7 resources)
− Imputed rent for owner-occupied housing
− Consumption of fixed capital excluding CFC on owner-occupier housing (P51c1, S14)

This income control can be computed for the period 2000-2015.[3] On average, it corresponded to 67% of GDP, with a slight upward trend over time, reaching 71% in 2015. Since detailed national accounts are not yet available for 2016, we assume the same proportion of GDP as 2015 to obtain the income control total for 2016. The figures for these years are

[3]This income control could also be computed for 1995-1999 but we opted not to due to the lower quality of the tabulations for these years. Thus, our income denominator relies solely on national accounts based either on the SNA-93 or the SNA-08. In any case, the results would not change significantly if we extended the series back to 1995.

certainly too low for the earlier years, so, instead of conditioning only on GDP, we exploit other macroeconomic aggregates available since 1947, like household consumption, direct taxes, valued added by the real estate sector and gross savings. More specifically, we make use of accounting identities to rewrite the formula for the income denominator as:

Total pre-tax fiscal income (1947-1994)
= Household final consumption
+ Gross savings
+ Income and wealth taxes paid by households
− Imputed rent for owner-occupied housing
− Consumption of fixed capital
− Residual

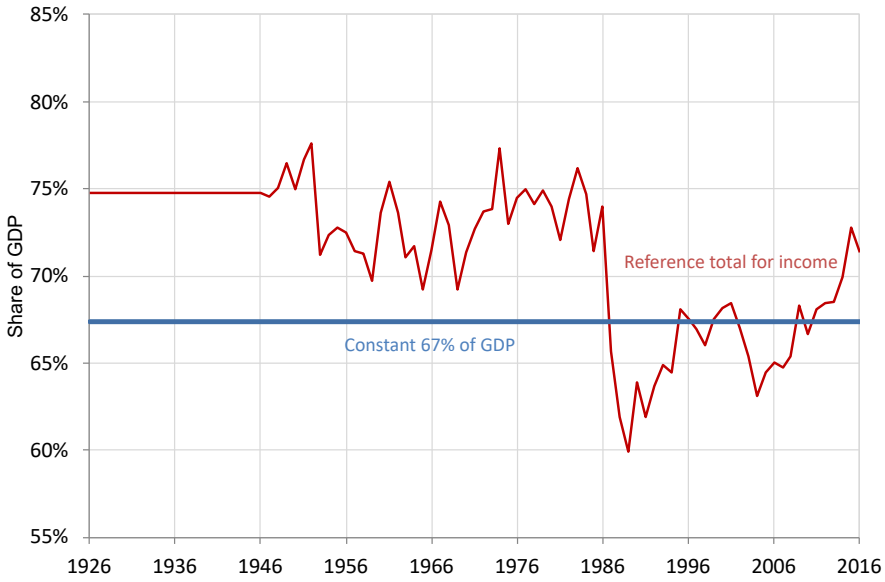
We compute for 2000-2015 the average ratio between each term and a similar macroeconomic aggregate available since 1947; namely, household final consumption expenditures, total revenue from direct taxes, value-added by the real estate sector and gross savings. The residual is anchored to GDP. These average ratios were used to calculate the control for total income between 1947 and 1999. For the earlier years (1926-1946), we use the same percentage of GDP as estimated for 1947.

Such a varying reference total for income is much more fine-grained than a constant percentage of GDP or any other national accounts aggregate because it accounts for the changing composition of the Brazilian economy. Figure B.4 contrasts our income denominator (as a fraction of GDP) with the one computed as a constant percentage. Our preferred, varying control for total income peaks at 77.6% and hits its floor at 60% of GDP (on average, it is about 72%). The constant percentage denominator clearly understates national income for most of the 20th century. In any case, one can easily show that if the reference total for income is multiplied by (1 + c), then the income share of the top x% drops by 1/(1 + c)%.

The integrated national accounts (*Contas Econômicas Integradas*, CEI) are available from the IBGE for the years 2000-2015 (IBGE, 2017). CEI follow the United Nations (UN, 2009) classification of institutional sectors and variables. All variables we use are sourced from the CEI, except for values of imputed rents, which we take from the IBGE's *Tabelas de Recursos e Usos* (TRU). Current Brazilian national accounts do not directly report the consumption of fixed capital (CFC) of households, so we take estimates of depreciation made by Morandi and Reis (2004) and extended by IPEA over the 1950-2008 period (about 11% of gross national income on average).[4] For 1947-1949 we use estimates of the share of depreciation in gross national income (about 5% of GNI) from FGV (1962), whose average we apply to the years 1926-1947. For 2009-2016, we assume the same proportion of CFC in GNI as 2008 (around 13%). In order to obtain fixed capital consumption for the household

[4]See `http://www.ipeadata.gov.br/Default.aspx`.

Figure B.4: Reference Total for Income vs Constant Fraction of GDP: Brazil 1926–2016



Notes: authors' calculations based on IBGE (2000, 2006, 2011, 2016b); IBGE,"Series Históricas"; Ipeadata; and Secretaria da Receita Federal. The constant 67% of GDP represents the average share for the 1995-2015 period, when more detailed national accounts data are available.

sector, we assume that it represents 25% of total CFC for the national economy, which is in-line with the ratios of Chile and Mexico, the only two Latin American countries with this data available on `wid.world`.[5] For 1995-1999, we use national accounts data published in IBGE (2000), while for the period 1947-1994, we use data from IBGE (2006), IBGE "Series Históricas", and Ipeadata.[6]

5.3 Generalized Pareto Interpolation

Since the income tax data are in the form of grouped tabulations, and since the income intervals do not generally coincide with the percentiles of the population with which we are concerned (such as the top 1%, the top 0.1%, etc.), we must resort to using an interpolation method to estimate the threshold values and average income levels for each fractile we are interested in, as well as to calculate the desired shares of total income. We use the generalized Pareto interpolation (gpinter) method of Blanchet et al. (2017) to turn the tabulated distribution into a continuous distribution for the portion of the population

[5]Government CFC and NPISH CFC are both set equal to the gross operating surplus of each sector in the national accounts, which amounts to about 11% of total CFC for the former and about 1% for the latter. Thus, we estimate corporate CFC as a residual, which comes to a share of about 63% of total CFC.

[6]See `seculoxx.ibge.gov.br`, `seriesestatisticas.ibge.gov.br`, and `ipeadata.gov.br`.

included in the tabulation.[7] The interpolation makes use of the population denominator (defined in section 5.1) to estimate fracile thresholds and average incomes, and the income denominator (defined in section 5.2) to estimate income shares. We apply gpinter to the tabulations whose thresholds and bracket averages are reported in the same income concept.[8]

This interpolation method, contrary to the standard Pareto interpolation method, allows us to recover an income distribution without the need for parametric approximations. It estimates a full "generalized Pareto curve" b(p) (i.e. a non-parametric curve of Pareto coefficients) by using a given number of empirical thresholds pi provided by tabulated data. In this way, the Pareto distribution is given a flexible form, which overcomes the constancy condition of standard power laws (with distributions being characterized by single Pareto coefficients), and produces smoother and more precise estimates of the distribution. In standard practice, the inverted Pareto coefficient b(p)=$\mathbb{E}(y|y > y(p)]/y(p)$ (where y(p) is the income or wealth quantile threshold corresponding to fractile p) is assumed to be constant within the top 10% or top 1% of the distribution. This approximation is acceptable for some purposes but it is generally not entirely correct. Moreover, it cannot be used to interpolate the portions of the distribution outside the top (see Blanchet et al. (2017)).

Using gpinter to individualize the distribution Income tax tabulations often refer to tax units, which corresponds either households or married couples. Thus, using tax units as the statistical unit can make results overly dependent on the demographic structure of the population, so it is preferable to "individualize" the distribution. The generalized Pareto interpolation (gpinter) package can do so under the convention that income is split equally between spouses. To do so, we need the fraction of singles (or equivalently couples) for all brackets. The method assumes that within each bracket, the income of singles is the same as the income of couples. This is almost certainly an approximation, but in practice it works well because the income difference between singles and couples is overwhelmingly a between-bracket phenomenon and not a within-bracket phenomenon. Therefore, the interpolation method we apply allows us to estimate the distribution of equal-split adult income from the tabulations. It must be noted that our equal=split series is not perfect, given that we do not know whether all single declarations that we estimate (see section 4.2) are made by individuals whose actual civil status is single, or whether they are from married persons filing separately from their spouses. Therefore, our procedure implicitly treats individually-filed declarations as being from single individuals, or from married individuals whose partners' income belongs to the same bracket. This implies that

[7] See `wid.world/gpinter/` for an online interface and an R package to apply the method.

[8] For the tabulations ranked by total income, we apply the interpolation to the total income variable in the tabulation; for tabulations ranked by gross taxable incomes, we apply the interpolation to the gross taxable income variable, and for tabulations ranked by net taxable incomes, we apply the interpolation to the gross taxable income variable.

our equal-split series may over-state inequality with respect to the perfect equal-split case (where all couples can be identified and have their income divided by 2), while under-state inequality compared to the pure individualistic case (where each spouse is assigned his/her own income).

Share-based interpolation As described in Section 4, we do not have tabulations for all years with reported incomes ranked by the same income concept we are interested in. Our broad strategy is to perform the interpolation, where possible, on the same concept incomes are ranked by in the tabulations. When the original ranking variable of a tabulation does not correspond to the income concept we wish to capture (total income; where missing, gross taxable income; where missing, net taxable income) we must pursue a different strategy for interpolating the income variable of interest. We resort to a share-based interpolation using the gpinter program. This variant of the method is tailored for situations where bracket-level information on the number of declarations and income declared exists without their threshold counterparts. It operates by interpolating the Lorenz curve with the information provided on population fractiles and their average income. This procedure is used to interpolate gross taxable income shares from tabulations ranked according to net taxable incomes (for 1926-1968), and interpolating total income shares from tabulations ranked by gross taxable incomes (for 1974-1988) and tabulations ranked by non-taxable incomes (for 1979-1988).

5.4 Constructing our Preferred Series for Top Shares from Tax Data and External Controls

Our preferred series comes from joining together the most complete estimates of top shares where we associate incomes from the different tax tabulations to reference totals from national accounts and demographic data (see sections 5.1 and 5.2). The rule of thumb we follow is to perform the interpolation for tabulations whose reported income was ranked by the same income concept, prioritizing the most complete income type (total income is preferred to gross taxable income, which is preferred to net taxable income). Where the ranking concept was distinct from the reported income of interest, we resorted to using share-based interpolations, as explained above.

At the end of the interpolation procedures, we have gross taxable income shares for the entire period. For 1974-2016 we have total income shares, with few exceptions, as obtained from the share-based interpolation for 1979-1988 and the generalized Pareto interpolation based on a complete tabulation for 2006-2016. Our preferred benchmark series is based on the total income series. There are two modifications we make to the post-1974 series for total income. First, for the years 1974-1988, our preferred Top 10% income share comes from interpolating total income shares from tabulations ranked by gross taxable income shares. For the Top 1% (and higher fractiles) our preferred series is based on the

interpolation of total income shares from the tabulations ranked by non-taxable income, available for the years 1979-1988. For the years 1974-1978 we re-scale the Top 1% share (and higher fractiles) based on the 1979 ratio between the share from the tabulation ranked by non-taxable income and the share from the tabulation ranked by gross taxable income. The reason for doing so is that the ranking of non-taxable income better reflects the concentration of total income than the ranking of gross taxable income, at least for the highest incomes, which we assume apply to the Top 1% and above. In 1979 and 1980 the difference is relatively small compared to the discrepancy that emerges between the Top 1% series of total income ranked by gross taxable income and that ranked by nontaxable income from the early 1980s. This widening gap is due a large portion of non-taxable incomes (mainly related to financial contracts) that through legislation benefited from close-to-perfect indexation to inflation (Kane and Morisset, 1993; Nóbrega, 2014).

The second modification we make is to exclude from our preferred series shares derived from the 1988 tabulation ranked by non-taxable incomes, whose levels we find difficult to justify. Annual inflation rates trebled from 1987 to 1988, but the reported incomes in tax data in 1988 change in an overly dramatic manner. Bracket average incomes explode, while the shape of the tabulated distribution changes completely.[9] We do so despite upwardly adjusting the reference total from national accounts for the year 1988. The reason for performing this adjustment is that 1988 saw a significant discrepancy between inflation measured via the national accounts deflator and the consumer price index (INPC). This difference persisted more or less until 1994 (see Figure B.5). It is significant since incomes reported in tax returns that were subject to monetary correction (i.e. indexation) were anchored to the INPC, while national accounts were based on the GDP deflator. Since we only have distributional data for 1988 between these hyper-inflationary years, we only scale-up the 1988 reference total to account for the differences in the price indicators. For 1988 the income control increases by 59% (leading to a drop in top shares by 63%) compared to its previous value when we apply the formula to the income control of 1987:

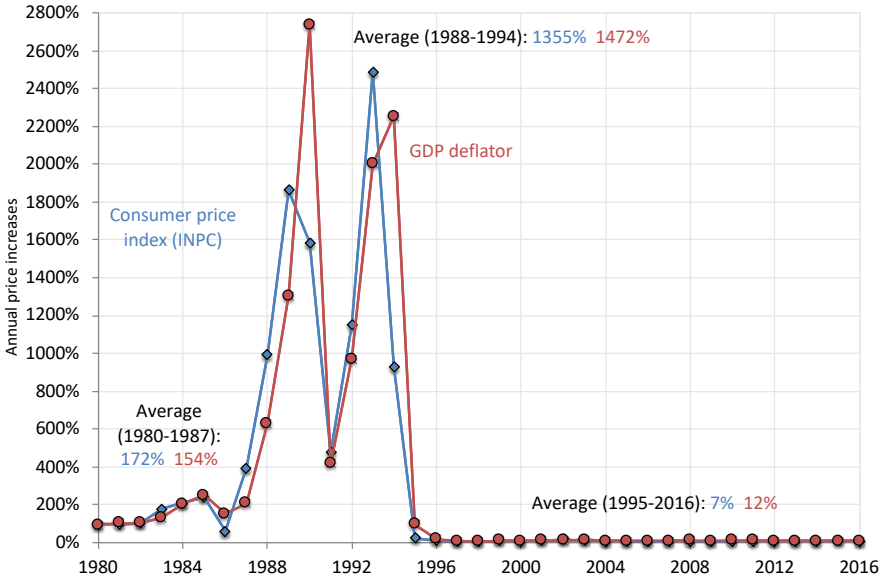
$$(1 + realgrowth_{1987-1988}) \times (1 + INPC_{1987-1988})$$

Figure B.6 depicts the results of our choices for our preferred series, alongside series derived from tabulations ranked by different income concepts for the period 1974-2016.

For the years 1926-1973, when we only have shares of gross taxable income, we must

[9]The 1987 and 1988 tabulations both have 18 income brackets. However, in 1987 the distribution of declarations per bracket resembles that of a "normal" distribution, with most of the mass located around the middle, while in 1988 the distribution is completely right-skewed. Moreover, while the top open bracket in 1987 contains 80,000, relating to the top 0.1% of the population declarations, the top bracket in 1988 only contains 31 declarations, which relates to the top 0.00004%. Moreover, the this top bracket in 1988 accounts for a disproportionate share of total reported income at 7%. It is hard to believe that incomes benefiting from monetary correction are driving the changes when the number of returns at the top changes so much. Over one year monetary correction should equally benefit those holding similar income sources.

Figure B.5: Inflation in Brazil according to GDP Deflator and INPC: 1980–2016



Notes: GDP deflator and the national consumer price index (INPC) are both source from the IBGE. We don't consider 1988 for our preferred series (see text for details).

rely on information for 1974-1987. Fortunately, the relationship between total income shares and gross taxable income shares was very stable during this period. Thus, our strategy for estimating total income shares for the earlier years is to add to gross taxable income shares the average value for each top fractile between 1974-1979, before the years of accelerating high inflation. In other words, the total income share of the Top 1% for 1926-1973 is the gross taxable share +12.48 percentage points (pp) for the equal split series and +12.40 pp for the individual series.

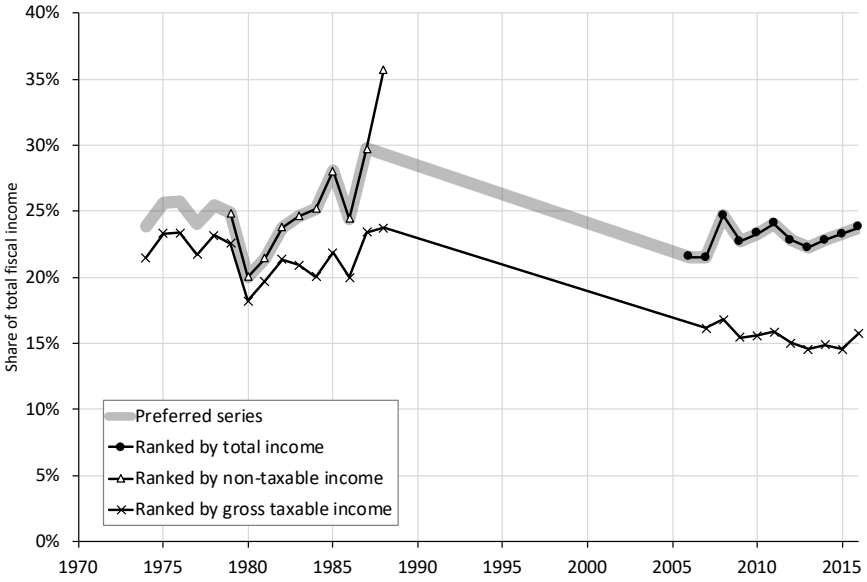
For the years 1996-1998, 2000 and 2002, the procedure to impute total income shares was different, because we could also rely on aggregate estimates for non-taxable incomes. Thus, we first run fractional response models (Papke and Wooldridge, 1996, 2008) for 2007–2016 and estimate predicted total income shares by percentile for 1996–1998, 2000 and 2002. Then we use these income shares to compute average incomes and adjust the thresholds accordingly, once again using 2007–2016 as reference.

More specifically, we run a fractional probit, that is, we assume that for all i:

$$E(y_i|\mathbf{x}_i = \Phi(\mathbf{x}_i\beta)$$

where y_i is the total income share of percentile i in 2007–2016, $\mathbf{x}_i$ is a vector of covariates, and $\Phi(.)$ is the standard normal CDF. The quasi-maximum likelihood estimator of β is obtained by $max_b \sum_{i=1}^{N} l_i(\mathbf{b})$, where the log-likelihood function $l_i(\mathbf{b})$ is given by:

Figure B.6: Top 1% Total Income Shares according to Tabulations Ranked by Different Income Concepts in Brazil: 1974 - 2016



Notes: authors' calculations based on data from tax tabulations and national accounts.

$$l_i(\mathbf{b}) \equiv y_i ln[\Phi(\mathbf{x}_i\mathbf{b})] + (1 - y_i) ln[1 - \Phi(\mathbf{x}_i\mathbf{b})]$$

The vector of covariates $\mathbf{x}_i$ includes a constant and two variables: the gross taxable income share of percentile i ($gross_i$) and the aggregate share of non-taxable incomes in total income in each year k ($nont_k$).

For the latter variable, there is a spike in the share of non-taxable incomes in 1997 caused either by a typo in the available tabulations or by an artificial one-off reaction to the law that exempted profits and dividends from personal income taxation. Indeed, while non-taxable incomes accounted for 29% of total income reported on tax returns in 1996, they soared to almost 48% in 1997 and quickly plunged back to the same level as before in 1998 (Souza, 2016, p. 176–177). Hence, we "smooth" $nont_k$ for 1997 by replacing the observed value by the average percentage from 1996 and 1998 in order to avoid introducing biases in our estimates

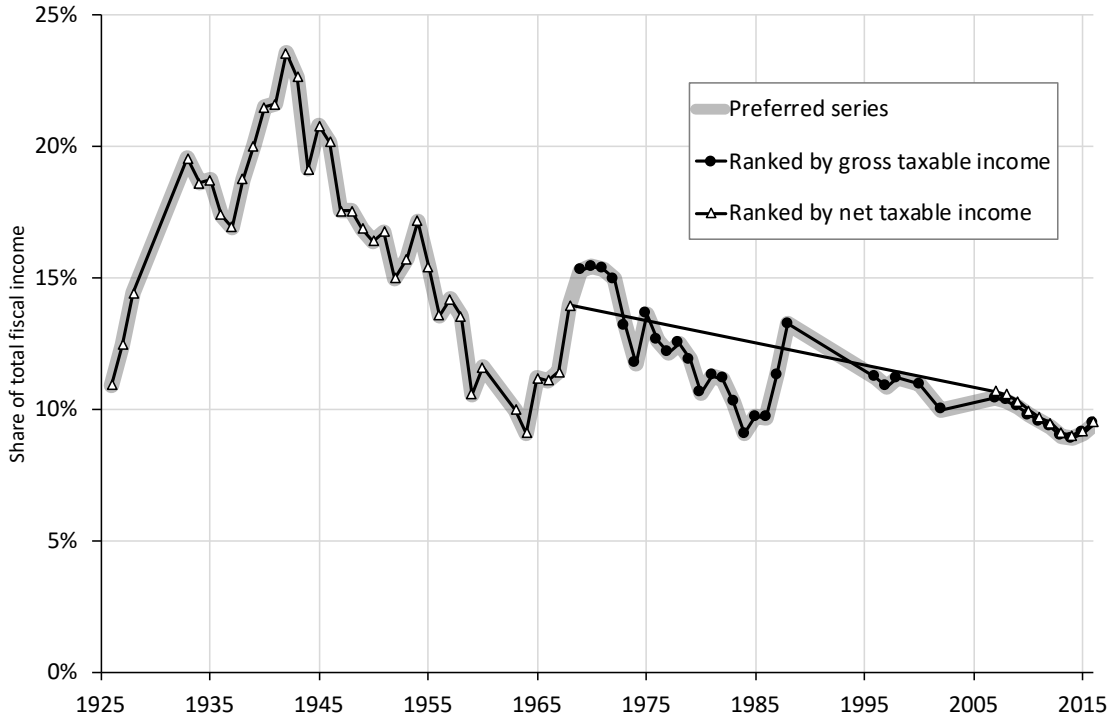
The fractional probit guarantees estimated income shares fall in the appropriate $(0, 1)$ range and is flexible enough to provide consistent point estimates even if the true underlying model is not a probit. More importantly, the comparison of predicted and observed values for 2007–2016 is very accurate.

Once we applied the coefficients estimated for 2007–2016 to the years with missing total income data (1996–1998, 2000, and 2002) it is straightforward to calculate all other quantities of interest, except for the adjusted income thresholds for each percentile. In this

case, we relied again on the years with complete information and calculated the average ratio between each percentile's income threshold and average income. Then, we applied these percentile-specific ratios to the years with missing data.

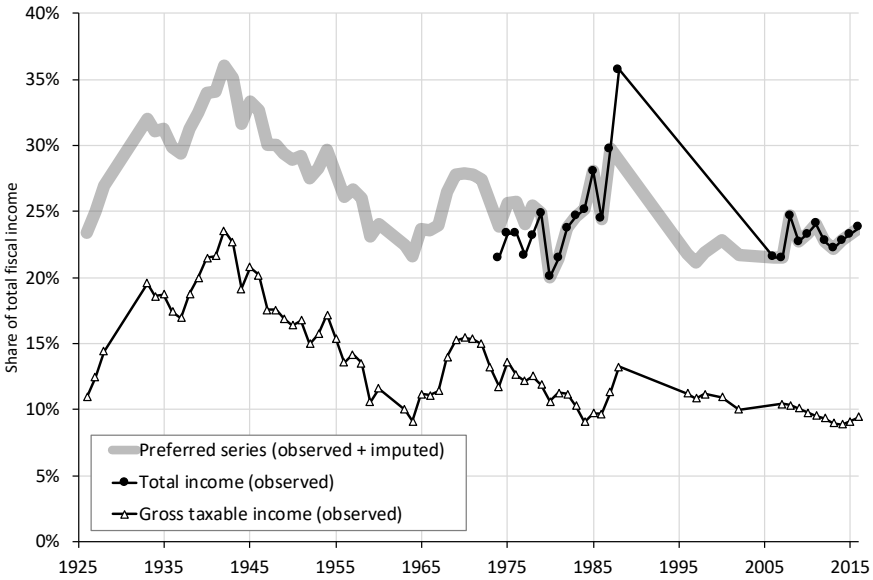
Figure B.7 presents the evolution of Top 1% gross taxable income shares according to tabulations ranked by different taxable income concepts, alongside our preferred series for gross taxable income. From our preferred gross taxable income series we arrive at the estimation of total income for the entire period, shown in Figure B.8. Our preferred series exploits the observed (non-imputed) series depicted in Figure B.6 and Figure B.7 and relies on an imputation (anchored to our preferred gross taxable income series) to reach 1926. It can be seen that our imputations mainly change levels rather than trends. On average, the total and taxable series follow similar trends during the 1970s and early 1980s, before the large inflation hike. Therefore, this stability provides us with the only likely trend of total income for the earlier years. Overall there seems to be no sharp discontinuities in our preferred series across different tabulated rankings, with the exception of the period from the mid-1980s, due to rampant inflation, and more recent years, due to changes in tax laws.

Figure B.7: Top 1% Gross Taxable Income Shares in Preferred Series and according to Tabulations Ranked by Different Income Concepts in Brazil: 1926 - 2016



Notes: authors' calculations based on data from tax tabulations and national accounts.

Figure B.8: Top 1% Income shares in the Preferred and Non-Imputed Series in Brazil: 1926 - 2016



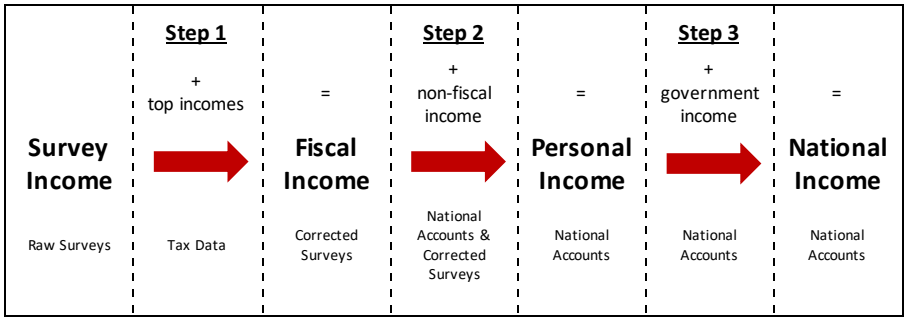
Notes: authors' calculations based on data from tax tabulations and national accounts.

6 Estimating DINA: 1976-2016

Figure B.9 depicts the general framework that we apply to build Distributional National Accounts (DINA) for Brazil, which is divided up into three steps. Our aim is to distribute total net national income, as it appears in the national accounts, among adult individuals residing in households. Using the surveys, we can first estimate a distributional series of "survey income". Combining the survey data with tax data, we can then compute a series of "fiscal income". The income captured in these two series covers pre-tax labour income, mixed income and capital income. More precisely this includes wages and salaries, pensions, self-employment income, interests, rents, distributed business profits and dividends, and capital gains. It thus corresponds to pre-tax post-replacement fiscal income, i.e. income received by individuals before personal income taxes, employee and self-employed social contributions (but after employer contributions), and legal deductions, but after accounting for social security benefits in cash (unemployment insurance and social security pensions). All these items are included or excluded in order to make the income in the survey consistent with the definition of income in the personal income tax declarations.

"Fiscal income" is distinguishable from "national income" insofar as it only concerns distributed income received by physical persons that is assessed by the tax office for fiscal purposes. It should also be distinguished from "taxable income", which is the income that is ultimately taxed after the application of legal deductions. Some components of income can be reported on the tax returns but are not taxable. This may vary with countries.

Figure B.9: Framework for Building Pre-Tax DINA for Brazil



Notes: authors' elaboration. Each income concept is associated to an underlying data source for its construction. See text for definitions.

As we have seen, in the case of Brazil it is explicit, as the tax declarations include a section for declaring non-taxable incomes. This fiscal income concept also excludes business expenses of independent workers required to keep accountancy books (e.g. doctors, dentists, psychologists, lawyers, independent commercial agents, etc.), as these expenses are incurred to generate their income. These expenses are can be identified in the deduction "livro caixa" in the tabulations, which we use to subtract from total assessed income. Such expenses are not identifiable in the household survey, but we know from the tax statistics that these generally affect higher incomes more, which the tax data does better to capture. Pre-tax fiscal income corresponds to the national accounts total we defined in section 5.2.

Moving from fiscal income to "personal income" implies that we factor in flows of income appearing in national accounts that that (1) get attributed to households but are not included in fiscal income, such as imputed rents, investment income attributable to insurance and pension funds and social contributions made by employees and self-employed workers; and (2) do not end up in households, but rather in corporations, such as undistributed corporate profits (i.e. net primary income of corporations). We also must subtract the social contributions made by employees and self-employed workers. Finally, moving from personal income to "national income" means that we must account for incomes that flow to the government rather than to households, namely government net capital income (government share of undistributed corporate profits, net interests and other incomes) and net production taxes. Thus, total pre-tax national income is computed as follows:

Total pre-tax national income (DINA)
= Total pre-tax fiscal income
− Social contributions (D61, S14)
+ Imputed rent for owner-occupiers
+ Investment income attributable to insurance policyholders (D441, S14)
+ Investment income payable to pension entitlements (D442, S14)

+ Household/NPISH component of pre-tax undistributed corporate profits (B5n, S11+S12)
+ Government factor income[10]
+ Pension and other social insurance surplus[11]
+ Net production taxes received by the government (D2-D3, S13)

In the following subsections, we detail how we move through the three steps depicted in Figure B.9.

6.1 From Survey Income to Fiscal Income

Step 1. In the first step, we start from the largest distributional data source at our disposal, the representative household survey (PNAD). We preserve the income sources that are known to be reported in our second distributional dataset, the personal income tax tabulations (DIRPF). We compare income levels in the tax data to those in the survey data to verify the extent of income under-coverage in the latter and incorporate the tax information into the survey data using a combination method developed in chapter 1. Using the distribution of incomes from the tax data, this method reweights survey observations within the original support (to address non-sampling error) and creates new observations at the top, matching survey covariates (using a nearest-neighbour imputation method) to the newly generated distribution (to address sampling error). This method confirms that substantial income under-coverage occurs at the top of the distribution (see Figure B.10 for evidence on selected years).[12] At the end of the procedure we have a distributional series for total fiscal income (income that would be reported if everyone were to file a declaration) for the overlapping years of both datasets. For years where only one of the two datasets is available we proceed to extrapolate/interpolate the estimates according to information from the closest relevant years with both sources.[13]

6.2 From Fiscal Income to Personal Income

Step 2. In this step, we incorporate into the fiscal income distribution all private non-fiscal incomes that can be attributed to the personal (household) sector, in order to arrive

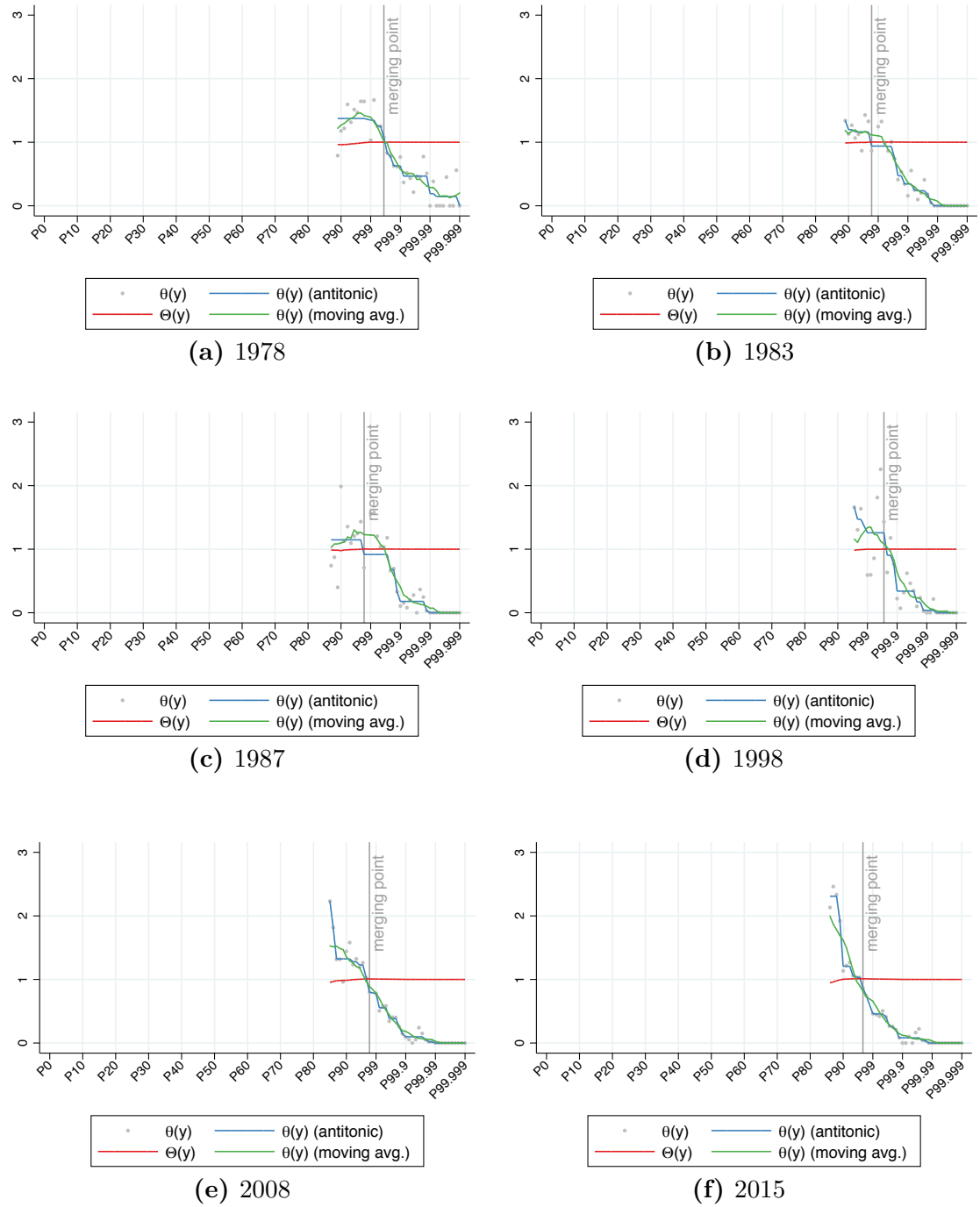
[10]This equals net property income received by the government (D4n, S13).

[11]This equals pension and other social contributions (D61, S14) − pension and other social insurance benefits (D621+D622, S14).

[12]On average, over the whole period, the "merging point" between the two datasets is found around percentile P98 in the distribution. We correct about 0.5% of the total population, with 94% of the corrected population being inside the survey support, and 6% being outside the survey support, i.e. above the maximum income observed in the survey.

[13]Table B.2 presents the strategy we employ for each year where we make these imputations during our estimation procedure. The one exception is our treatment of 1988. For reasons highlighted in section 5.4 we choose to ignore the income information from the tax data for this year for our benchmark estimates and anchor estimates from 1988-1993 to the 1987 relation between survey estimates and joint survey + tax data estimates.

Figure B.10: Merging Point between Surveys and Tax Data in Brazil for Selected Years



(a) 1978 **(b)** 1983

(c) 1987 **(d)** 1998

(e) 2008 **(f)** 2015

Notes: the panels in the Figure depict the estimated bias in the surveys (PNAD) relative to the tax data (IRPF). $\theta(y)$ (grey dots) are the ratio of income density in the survey over that of tax data for each quantile of the fiscal income distribution. The green line is the centered average of $\theta(y)$ at each quantile and eight neighboring estimates. The blue line is the result of an *antitonic* regression applied to $\theta(y)$. It is constrained to be decreasing in order to find a single merging point. The red line is the ratio of the cumulative densities. See Blanchet, Flores and Morgan (2018) for further details.

at a distributional series for pre-tax post-replacement personal income. This requires the identification and imputation of missing personal capital income as well as to the imputation of social contributions, which are not included in the gross income assessed for fiscal purposes. The missing personal capital income is income attributed to households but not declared to the tax authorities, and also income that does not get attributed to individuals, but rather to corporations. The first part we can identify as investment income attributable to pension and insurance funds held by individuals and imputed rents, while the latter are the undistributed profits of privately owned corporations associate to Brazilian residents. Thus, in this step we impute the following items:

Social insurance contributions (D61, S14)
Imputed rent for owner-occupiers
Investment income attributable to insurance policyholders (D441, S14)
Investment income payable to pension entitlements (D442, S14)
Household/NPISH component of pre-tax undistributed corporate profits (B5n, S11+S12)

The totals of each of these items are either retrieved directly from national accounts or estimated if the accounts are not sufficiently detailed (this is the case generally for years before 1995). For social contributions, we directly retrieve the totals from the IBGE's system of national accounts 1995-2015 IBGE (2000; 2017). The pre-1995 totals are estimated as a residual from the share of total direct tax revenue in GDP (IBGE, 2006) and the shares of personal income, property and corporate income taxes in GDP (IBGE, 2006).[14] Overall, social contributions represent an average 9% of national income over 1976-1994 and 14% over 1995-2015. Totals from imputed rent are obtained from IBGE (2017) for the 2000-2015 period. Before 2000, we use estimates of imputed rent made by Souza (2015; 2016), who computed the average ratio between imputed rent and the value-added by the real estate sector over the 1995-2014 period and applied it to values of the latter variable, available since 1947.[15] On average, this imputed rent makes up about 7% of national income between 1976 and 1994 and 9% between 1995 and 2015. For investment income on private insurance and pension funds we source the aggregates for 1995-2015 from IBGE (2000; 2017).[16] For previous years we follow two different estimation strategies for each of the two components. For D441, we apply the average share in GDP between 1995 and 2015 (0.08%) to all years between 1976 and 1994. For D442, we make use

[14]For years before 1995, the share of these three taxes in GDP is estimated back using the growth rate of (total income tax revenue/GDP) from IBGE (2006).

[15]The only modification we make to this strategy is that by using the latest SNA for the 2000-2015 period, the estimates of imputed rent are slightly distinct from the estimates used by Souza (2015; 2016). Thus, we anchor his pre-2000 estimates the latest series available in IBGE (2017) by multiplying each year by the difference in the year 2000.

[16]Prior to 2010, the national accounts did not split D44 into three separate categories (D441, D442 and D443). Thus, we assume the same proportions for each category in the available D44 aggregate over 1995-2009.

of data on the total assets managed by private pension funds (*previdencia complementar fechada e aberta*).[17] We performed an OLS regression of the D442/GDP ratio on the ratio of total private pension fund assets/GDP to obtain predicted values for the 1985-1994 period. We extrapolate the 1985 share (0.42%) back to 1976. The total fund income we seek to distribute amounts to about 0.7% of national income over 1976-1994 and 1.2% over 1995-2015.

The final aggregate in this stage of the estimation process concerns pre-tax undistributed corporate profits owned by the domestic Brazilian households. This corresponds to the net primary income of the corporate sector (B5n) attributable to the domestic household sector (S14). In order to estimate this series we proceed through various steps. In the first step we estimate a series of gross primary income of the corporate sector back to 1976. For 1995-2015 we avail of data from IBGE (2000; 2017) on the gross primary income of the corporate sector. For the years 1976-1994, we extrapolate backwards using the growth rate of the sum of gross corporate savings and the corporate income tax. Figures for both for the period 1995-2015 are obtained from IBGE (2000; 2017). For 1976-1994, we apply the growth rate of gross private savings (IBGE, 2006) to gross corporate savings in 1995, and apply the growth rate of income tax (from IBGE (2006)) to the values of corporate income tax in 1995. The second step involves deducting corporate depreciation to arrive at net primary corporate income (i.e. pre-tax undistributed profits). As explained in section 5.2, we have estimates of corporate depreciation for the period 2000-2015. For all prior years we assume the same share of corporate depreciation in total national depreciation observed in 2000 (61%). The third step requires us to estimate the household share of the net primary corporate income. For this, we make use of the financial accounts from 2010-2015 (IBGE, 2017) to compute the household share of total corporate equity liabilities (AF5), as separate from the government and foreign sector shares in total equity. We derive this share as the residual from the share of government and foreign net equity assets in total corporate equity liabilities. The household share is on average 59%. For previous years, we make use of proxy data from two business surveys conducted by *Visão* magazine ("Quem e Quem na Economia Brasileira") in 1974 and 1985 on the 5,113 and 8,094 largest incorporated firms in the country in each respective year. These surveys collected data on firms' sales revenue, profitability and assets, among other information. According to the data, in 1974, 48% of total net assets belonged private domestic firms, 15% to foreign multinationals and

[17] Previdencia "fechada" (closed insurance) is a private insurance system sponsored by firms, whereby employers match funds contributed by their employees. This system dates back to the 1960s, starting with large state-owned companies. The previdencia "aberta" (open insurance) is a more recent phenomenon of the 1990s, whereby individuals make their own deposits into a managed fund, whose balance grows over time with contributions and investment income. For closed funds we use data from http://www.abrapp.org.br/Consolidados/Consolidado_Estatistico_1997_12.pdf,and http://www.abrapp.org.br/Consolidados/Consolidado_Estatistico_2004_12.pdf, while for open funds the data come from http://repositorio.unb.br/bitstream/10482/16437/1/2013_JulioCesarAlvesVieira.pdf.

37% to public enterprises. In 1985, these shares were 43%, 9% and 48% respectively.[18] We approximate the household share in corporate assets (and thus undistributed profits) from the private sector control of total assets from these surveys (48% in 1974 and 43% in 1985). We linearly interpolate the shares for years between the estimates we have for 1974, 1985 and 2010. Multiplying these shares to our net corporate primary income series gives us an annual series for the household component of pre-tax undistributed profits. In total it represents about 5% of national income for the 1976-1994 period and about 6% for the 1995-2015 period.

At this stage, we have the aggregates for each of the non-fiscal income categories we need to impute to our distribution. Thus, the second ingredient we require is their distribution among Brazilian individuals. We impute the distribution of all four categories using the PNAD survey. Before its correction we have already imputed values for imputed rent per adult and social contributions per adult (see section 4.3). After we have corrected the survey using tax data (6.1), we impute the aggregates of the other remaining categories – insurance and pension fund investment income and pre-tax undistributed corporate profits associated to households. For fund income, we distribute the aggregated total (D441+D442) by assuming that it follows the distribution of primary job income for earners who contribute to a social insurance fund (i.e. who have positive social contributions).

For undistributed corporate profits, we test eight imputation scenarios, which can be grouped into four broader scenarios.[19] Scenario 1 assumes that these profits follow the joint distribution of financial income and employer profit withdrawals. Scenario 2 assumes that they follow the distribution of total capital income (financial income, rent and employer profit withdrawals). Scenario 3 assumes that they follow the joint distribution of financial income and rent. And Scenario 4 assumes that they follow the distribution of financial income. Scenarios 1 and 2 are each split into three sub-scenarios, according to different assumptions regarding the estimation of employer profit withdrawals, which cannot be cleanly identified in the survey.[20] Sub-scenario A assumes that employer capital withdrawals are the portion of primary job income of employers that exceed the minimum wage; sub-scenario B assumes that they are the portion that exceed the exemption threshold for the personal income tax; and sub-scenario C assumes that they are the portion of job-related income of employers that exceed the salary ceiling for social contributions.

These different ways of allocating undistributed corporate profits among the population each reflect expectations we make on the 'beneficial owners' of them. Scenario 1 proxies these beneficial owners of profits by the individuals that receive interests, dividends and

[18] These estimates are reproduced from Bacha (1980) and Baer (2014).

[19] For the other non-fiscal income imputations, we do not test various scenarios for two main reasons. First, multiple scenarios can be less easily constructed given the auxiliary variables we use for them. And second, the assumptions we make on their distribution seem to be reasonable, given the incomes we are dealing with.

[20] The survey only reports the total income of employers, which conflates their wage and their capital-related withdrawals.

profit-related withdrawals from being an employer. If we think owning real estate is a necessary correlate of owning companies then Scenarios 2 and 3 should be considered; the latter being preferred if we disregard that employers withdraw profits that are not classified as dividends. Our preference is to base our imputation on the remuneration of employers (business owners) that can be attributed to profit withdrawals and the distribution of financial income, which is likely to be the best correlate of owning enterprises. Choosing to impute solely on the basis of the distribution of financial incomes (scenario 4) is likely overstate the concentration of corporate ownership, given that distributed profits tend in general to me more concentrated than total equity ownership (Alvaredo et al., 2017; **?**). On the other hand, imputing solely on the basis of employer capital withdrawals in the survey would likely understate the concentration of corporate ownership, given the absence of employers from the largest corporations from the sample. Our benchmark scenario is scenario 1B, where employer capital withdrawals are that portion of an employer's income that exceeds the exemption threshold for the personal income tax. This is motivated by the expectation that most employers would decide the level of their capital withdrawals as a residual of how much they want their labour income to be taxed or how much they desire a public pension in the future.[21] If employers want to pay as little tax as possible, without wishing to maximise their future public pension (given that they have private option, for instance), then they would pay themselves a salary that is just below the exemption threshold in the personal income tax schedule, and contribute the corresponding amount to social security. If they care relatively little about receiving a public pension in the future, then they would pay themselves a minimum wage salary to contribute the minimum to social security. If on the other hand, they place a high weight on maximising their future public pension, they would pay themselves the salary that would meet the contributory ceiling. Thus, the remuneration decisions of employers in Brazil depends, monetarily, on balancing these two competing desires - the amount of tax to pay in the present and for the future level of public pension to be received out of their current contributing salary. Our benchmark choice in this part of the allocation is justified by it being a middle-ground in this trade-off.

Overall, our benchmark scenario (scenario 1B) is justified by what we deem to be the most reasonable allocation of undistributed profits, given the variables at our disposal and our knowledge of the tax and social contribution system. This choice is further validated by the sensitivity analysis we depict in section 6.4.

[21] In Brazil distributed profits are not subject to the personal income tax (unlike most countries) nor the schedule for social contributions (like most countries). Public pensions depend on the contributions made out of the contributing salary of each worker.

6.3 From Personal Income to National Income

Step 3. Up until this step, we have imputed all the private non-fiscal incomes to the personal sector. The final step involves imputing the remaining categories of income to arrive at a national income distributional series. Here we must account for:

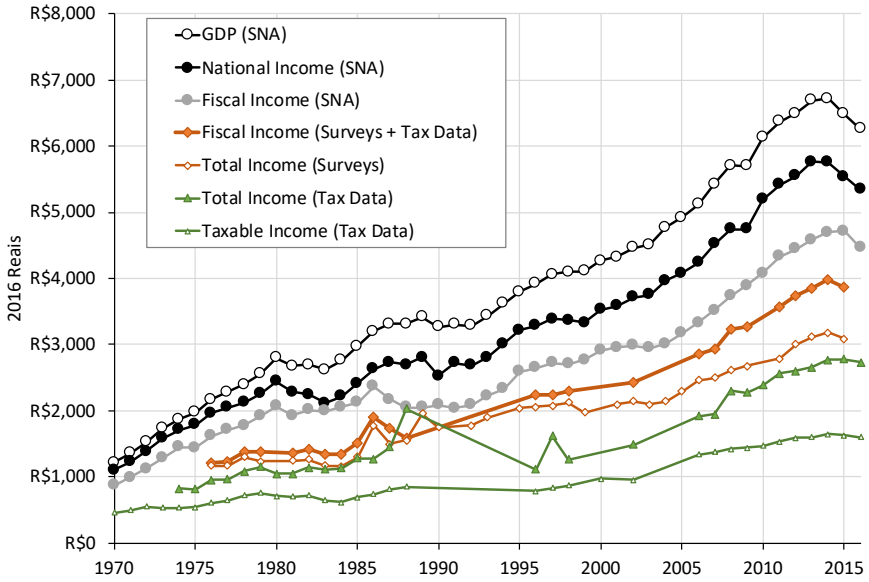
Government factor (capital) income (D4n, S13)
Net production taxes received by the government (D2-D3, S13)
Pension and other social insurance surplus (D61 – (D621+D622), S14)

We assume the distribution of these three categories is neutral, that is, we allocate it in a proportional way to the distributional of personal income, given a lack of a more precise method to impute each item. This choice can be motivated on the grounds that income from government assets should belong equally to all citizens. This reasoning can also apply to the final balance in social security funds, which has registered a slight deficit in Brazil over the last 20 years. One could argue that the deficit is covered by taxes/levies applied across the entire population on different sources of income/consumption. Net production taxes received by the government include value-added taxes and other product taxes, net of product subsidies, (D21-D31) and other taxes on production, net of subsidies (D29-D39). Along with government net property incomes, they form part of government's net primary income. In theory, these taxes should be imputed in relation to their incidence, which can be approximated by consumption, except for property taxes in D29, which should be imputed in proportion to housing wealth or imputed rent (Alvaredo et al., 2017). However, in practice imputations based on tax incidence are difficult to make without access to good data. Moreover, imputing these taxes in proportion to total personal income seams a reasonable first approximation, given that they are based on consumption-type taxes and to production/property-type taxes, which practically cover the entire population. Thus, in proceeding this way for these remaining items, we only seek to normalise the distribution to national income. This step, by definition has no impact on inequality.

Figure B.11 depicts the evolution of the main aggregate income concepts that we mobilise in the estimation process depicted in Figure B.9 from the survey data, the tax data and the system of national accounts (SNA). Through the estimation process just described, we arrive at the distribution of national income, which grew at 2.7% per year on average. It can be seen that other income aggregates grew at broadly comparable rates. Total (pre-tax) income in the surveys represent about 55% of national income, and approximately 72% of an equivalent income total from SNA. Total pre-tax income from tax data, while only covering 20% of the population in surveys, represents close to 50% of national income, and close to 60% of total fiscal income from SNA. Our merged tax+survey fiscal income series covers about 80% of an equivalent total from SNA and equals an average of 65% of total national income over the period. Thus, using this aggregate as a benchmark,

approximately 34% of national income is accounted for by "non-fiscal income", which we impute as previously outlined.

Figure B.11: Comparison of Income Aggregates by Income Source in Brazil: 1974-2016



Notes: authors' calculations based on data from surveys, tax tabulations and national accounts (SNA). Nominal values are deflated using the GDP deflator. National income covers about 85% of GDP. The remaining 15% is due to capital depreciation and net foreign income. Fiscal income (SNA), covering an average of 82% of national income, represents an equivalent income total to the income declared on tax returns, which itself accounts for close to 50% of national income. The equivalent income total in the raw surveys covers 55% of national income, while our corrected survey series covers close to 70% in recent years. This series merges total survey income with assessed incomes from tax data, of which about 60% comprises of taxable income and 40% of non-taxable income.

Two "anomalous" data points stand out in the graph. They refer to total income from tax data in 1988 and 1997. 1988 was the first year of the extreme inflation period, lasting until the end of 1994, that saw consumer prices treble from 394% in 1987 to 994% in 1988. Inflation would remain over 1000% on a yearly average up to 1994 (see Figure B.5). The highest incomes in tax data benefited from a monetary corrected that was quasi-perfectly indexed to inflation (for some sources, such as savings/fixed term accounts the correction was greater than inflation (Kane and Morisset, 1993)).[22] In spike in 1997 for tax data total income is less clear. It may relate to legislation passed in late 1995 that exempted distributed business profits and dividends from the personal income tax base (Law No 9.249 of 26/12/1995, Nóbrega (2014)). This can be verified by the fact that the increase in

[22]The monetary correction formula applied to personal incomes was based on the the consumer price index, which evolved in the same manner as the GDP deflator but registered different levels during the high inflation years (when prices were sometimes revised more than once per day) For our computations of inequality we ignore the information from tax data for 1988.

total income is due to the exceptional increase in non-taxable incomes in the tax statistics during this year, which dividends now comprised. The law came into effect in January 1996, but it is possible that behavioural responses from companies and shareholders took one year to adjust to the new changes. The spike may reveal that corporate owners distributed themselves the earnings accumulated in previous years. However, it is also possible that the amount of non-taxable income recorded in 1997 represents a typographical error in the publication, since no similar increase in incomes recorded in national accounts (dividends or total income) can be observed. In light of this uncertainty, we smooth the evolution of these incomes in the calculation of our benchmark fiscal income series (Surveys+Tax Data), as explained in section 5.4.

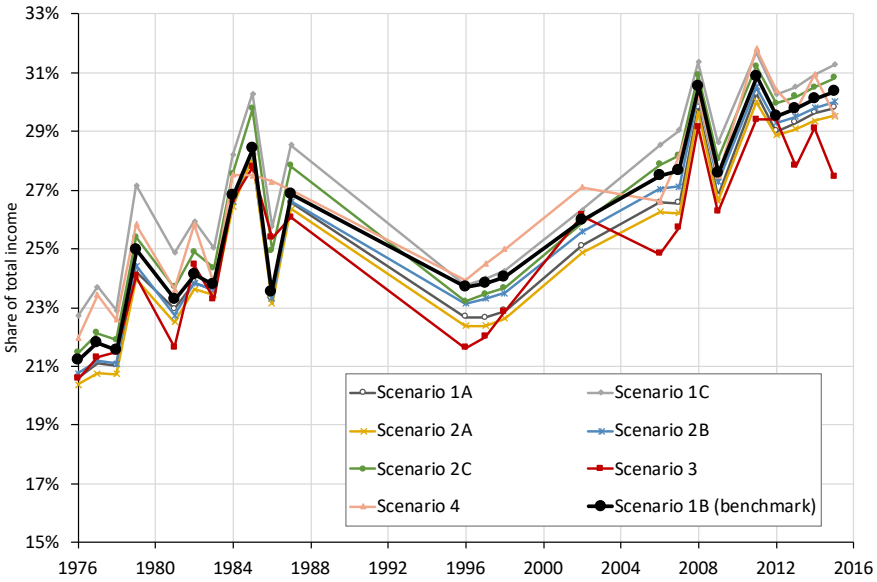
6.4 Sensitivity Analysis

Figure B.12 presents a graphical sensitivity test for the Top 1% income share in relation to the different imputation scenarios we perform for undistributed corporate profits (see section 6.2). The eight different scenarios for the the imputation of undistributed profits are based on different assumptions regarding their distribution across the population. It can be seen that our benchmark scenario represents a middle-bound estimate among the alternative scenarios that we test.

6.5 Filling in Missing Years

Between 1976 and 2016 we have intermittent coverage of distributional data from our two principal sources - the PNAD survey data and the DIRPF tax data. Step 1 of building DINA (see section 6.1) involves matching data from both of these sources, when they were both available for the same year. Thus, the resulting estimation of DINA, based on this method, produced results that only partially cover the last 40 years (in Table B.2., where both PNAD and IPRF equal 1). Our final benchmark series takes account of the data gaps, by incorporating estimates for years where at least one of the two distributional data sources is available. As a rule, we do not make estimates for years for which we do not avail of distributional data from either source. When either one of the two sources is available we proceed to impute the income shares from interpolating/extrapolating the remaining information from the most relevant neighbouring years. Table B.2 summarizes our strategy for these missing years. We follow two separate strategies depending on whether PNAD or IRPF data is missing. When we only avail of tax data (4 out of 41 years), we generally interpolate the top shares using the average of the shares of the closest neighbouring years, or using the difference between Tax+SNA estimates and Tax+Survey estimates, and impute bottom shares using the average composition of bottom shares from the neighbouring years. When we only have access to survey data, we extrapolate the shares based on the difference between Tax+Survey estimates and Survey estimates from the most relevant nearest years. For our national income series, we follow the same strategy,

Figure B.12: Comparison of Different Imputation Scenarios for Undistributed Profits on the Top 1% Share in Brazil: 1976–2016



Notes: Distribution of pre-tax national income to the Top 1%. Scenario 1 assumes that these profits follow the distribution of financial income (interests and dividends) and employer capital withdrawals. Scenario 2 uses the distribution of total capital incomes. Scenario 3 uses the joint distribution of financial income and property rent. Scenario 4 uses the distribution of financial income. Case A assumes that employer capital withdrawals are the portion of employer income that is greater than the minimum wage. Case B assumes that they are the portion greater than the annual exemption limit for the personal income tax. Case C assumes they are greater than the annual maximum contributing salary for social contributions. Authors' calculations based on data from surveys, tax tabulations and national accounts (SNA). We only show years for which a direct imputation was possible, i.e. years for which we have overlapping survey and tax data.

but base the extrapolations on the difference between the Tax+Survey+SNA estimates and the Tax+Survey estimates from neighbouring years.

Table B.2: Data Coverage and Estimation of Shares for Missing Years: 1976-2016)

Year	PNAD	IRPF	Estimation Strategy for Years with Data Gaps
1976	1	1	
1977	1	1	
1978	1	1	
1979	1	1	
1980	0	1	Interpolated using average of 1979 and 1981 shares.
1981	1	1	
1982	1	1	
1983	1	1	
1984	1	1	
1985	1	1	
1986	1	1	
1987	1	1	
1988	1	1	Information from tax data ignored. Results extrapolated using difference between Tax+Survey and Survey estimates from 1987.
1989	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 1987.
1990	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 1987.
1991	0	0	
1992	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 1987.
1993	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 1987.
1994	0	0	
1995	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 1996.
1996	1	1	
1997	1	1	
1998	1	1	
1999	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 1998.
2000	0	1	Interpolated using average of 1999 and 2001 shares.
2001	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 2002.
2002	1	1	
2003	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 2002.
2004	1	0	Interpolated using average difference of shares in 2003 and 2005 Between Tax+Survey and Survey estimates.
2005	1	0	Extrapolated using difference between Tax+Survey and Survey estimates from 2006.
2006	1	1	
2007	1	1	
2008	1	1	
2009	1	1	
2010	0	1	Interpolated using average difference in top shares in 2009 and 2011 between Tax+SNA and Tax+Survey estimates, and the average composition of bottom shares in 2009 and 2011.
2011	1	1	
2012	1	1	
2013	1	1	
2014	1	1	
2015	1	1	
2016	0	1	Interpolated using average difference in top shares in 2015 between Tax+SNA and Tax+Survey estimates, and the average composition of bottom shares in 2015.

Notes: PNAD is the household survey, IRPF is the personal income tax. SNA refers to System of National Accounts. 1 indicates data availability, 0 indicates no data availability. IRPF information is available for 1988, but we choose to ignore it for its lack of credibility.

7 Estimating DINA: 1926-1975

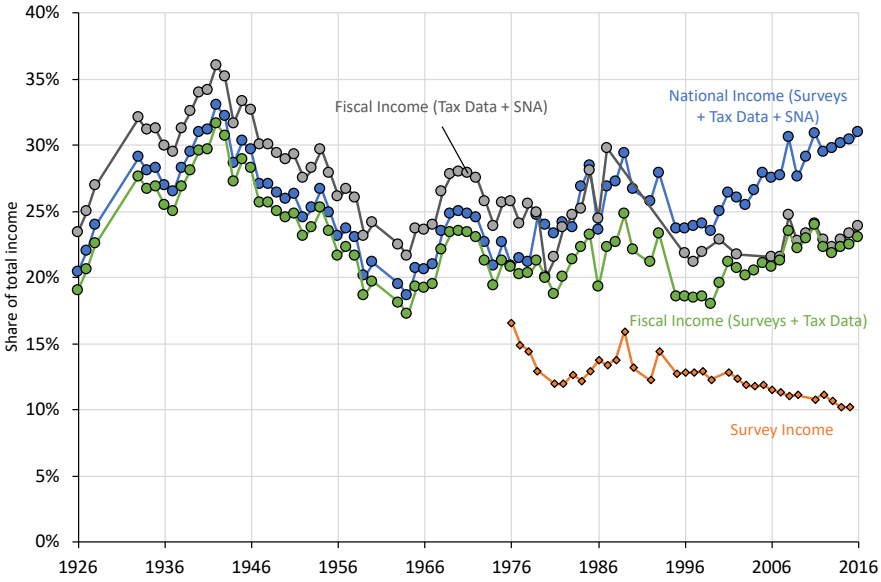
The previous steps detailed how we build DINA with all the data we have available for the period 1976-2016. In this section we describe how we proceed for the years prior to 1976, when we have no distributional information from surveys. We are thus left with limited coverage tax data and denominators from national accounts, alongside the post-1975 estimates. Our strategy to complete the fiscal income and national income series since 1926 is as follows.

As described in section 5 we compute a fiscal income series for top incomes using the tax data and reference totals from national accounts for the entire 1926-2016 period. The population coverage in the tax data allows us to estimate the Top 10% income share back to 1968. Before this we can only continue to estimate the Top 1% share (as well as shares of higher fractiles). In general, our strategy is to anchor our preferred fiscal income series (estimated from tax data and surveys for 1976-2016) and our national income series (estimated from tax data, surveys and remaining income from national accounts for 1976-2016) to the evolution of the fiscal income series estimated using tax data and an income denominator from the system of national accounts (SNA). More specifically, we extrapolate the fiscal income Tax+Survey series back to 1926 by adding the average difference in top shares between the Tax+Survey series and the Tax+SNA series for the years 1976-1979, to the Tax+SNA top shares for the preceding years.

Similarly, for the national income shares, we extrapolate the 1976-2016 series back to 1926 by adding the average difference in top shares between the fiscal income Tax+Survey series and the national income Tax+Survey+SNA series for the years 1976-1979, to the extrapolated fiscal income Tax+Survey shares for the years 1926-1975. Given that the survey is an integral part of the DINA estimation process there is little else we can do to estimate a distribution of national income prior to the survey's first data point. Thus, we implicitly assume that the average difference between the total fiscal income from SNA and the total fiscal income from a merged survey and tax data series over 1976-1979 is maintained over the entire period, rather than assuming that the "correct" income denominator prior to 1976 is given by our reference total estimate from SNA.[23] The relative stability in the difference between the series over the 1976-1979 period and relatively small average difference (about -1.0 percentage point for the Top 10% share and +1.8 points for the Top 1%) at least assures us that we are not making questionably large adjustments. Figure B.13 shows the discrepancy between all our final series. We dwell on the implications of these differences in section 4.1 in the main paper.

[23] As mentioned previously the quality/reliability of national accounts deteriorates the further we go back in time.

Figure B.13: Top 1% National and Fiscal Income Shares in Brazil by Concept and Source: 1926–2016

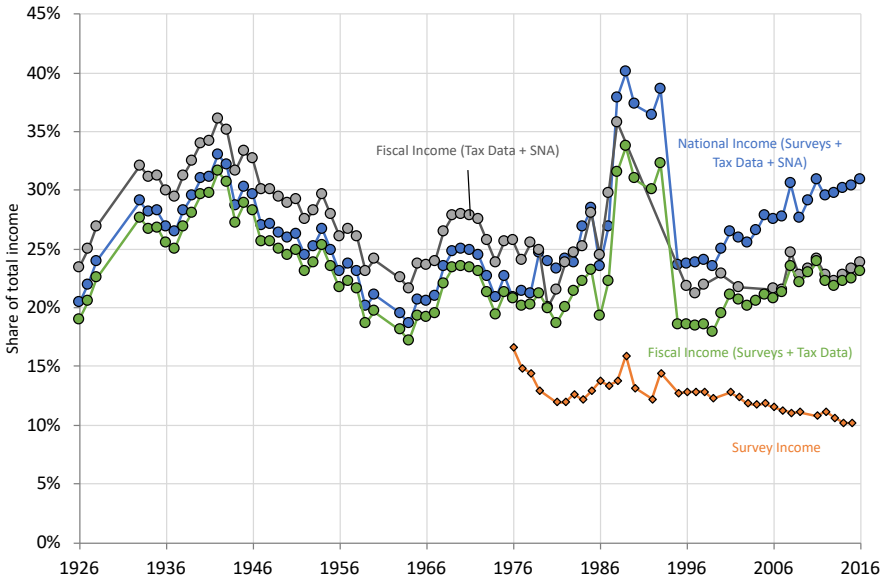


Notes: Distribution of pre-tax income to the Top 1%. Authors' calculations based on data from surveys, tax tabulations and national accounts (SNA). The national income series is expressed as a share of total national income. The fiscal income series are expressed as a share of total fiscal income. Our benchmark fiscal income series (Surveys + Tax Data) and National income series include our adjustments made to estimates between 1988-1993. Estimates for these years are anchored to 1987 estimates. See Table B.2 and section 5.4 for more details.

8 Top Shares using 1988 Income Tax Declaration

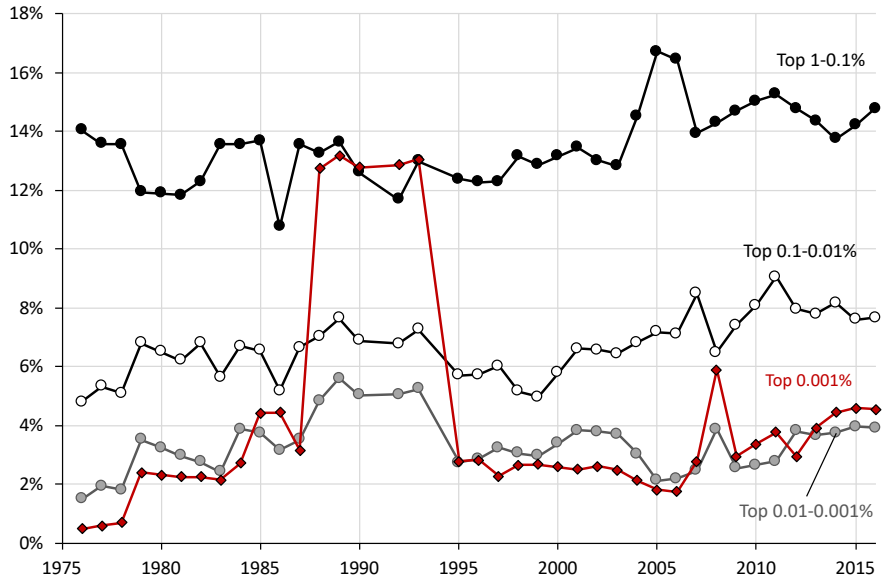
The series in Figure B.13 can be contrasted with those in Figure B.14. The latter include the raw estimates for 1988, based on the 1988 tax declaration, and subsequent estimates for the hyper-inflationary period (1989-1993) based on information from this year. Figure B.15 clearly shows that the remarkable jump in top shares is due to the very summit of the distribution (the Top 0.001%), who were declaring dis-proportionally high incomes in 1988 compared to 1987 (see section 5.4).

Figure B.14: Top 1% National and Fiscal Income Shares in Brazil by Concept and Source (including raw 1988 estimates): 1926–2016



Notes: Distribution of pre-tax income to the Top 1%. Authors' calculations based on data from surveys, tax tabulations and national accounts (SNA). The national income series is expressed as a share of total national income. The fiscal income series are expressed as a share of total fiscal income. The fiscal income series (Surveys + Tax Data) and National income series here include estimates for 1988-1993 based on information from the 1988 tax declaration. Thus, estimates for 1989-1993 are anchored to 1988 estimates.

Figure B.15: Income Shares in Brazil Between Top Groups: 1976–2016



Notes: Distribution of pre-tax national income among equal-split adults across income concepts. The unit is the adult individual (20-year-old and over; income of married couples is split into two). Fractiles are defined relative to the total number of adult individuals in the population. Authors' calculations (combining survey, tax and national accounts data). Estimates for 1988 use the raw information from the 1988 tax return, to which estimates for 1989-1993 are anchored.

9 Other Dimensions of Inequality

9.1 Gender Inequality

The division of labour income among the two sexes we depict in the paper mainly uses available information from tax data, alongside survey information. We explain in this section how we arrive at our estimates.

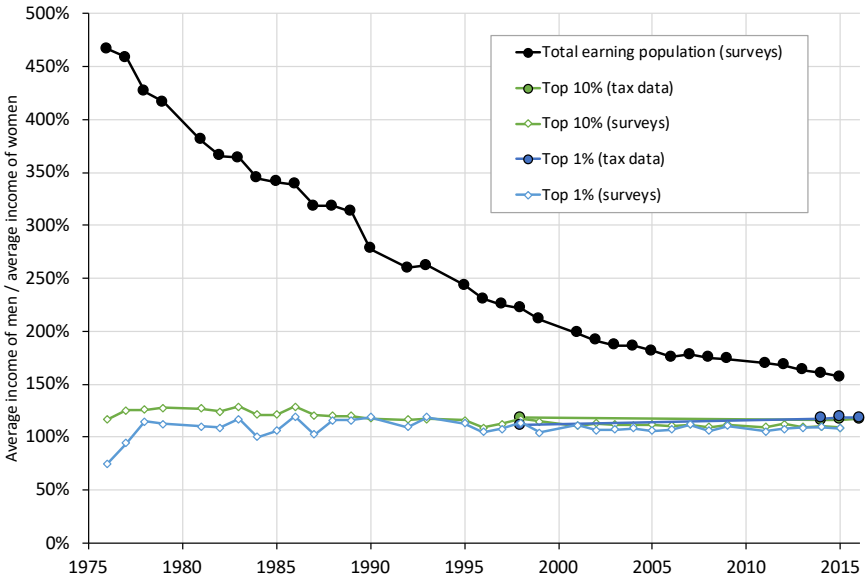
To divide labour earnings by gender, we use information on the composition of total "taxable income" from the tax data, available for the years 1974–1988, 1998 and 2014–2016. As explained in previous sections, this income concept mostly comprises of pre-tax labour income (wages of salaried and self-employed workers and pensions), but also property rent. This has been so since the tax reforms of the early 1990s. For the 1974–1988 period taxable income also included other capital incomes like interests and profits/dividends. Thus, to arrive at estimates for pure factor labour remuneration for the working-age population (20–64), we must exclude capital incomes and pensions. To do so, we apply ratios of (factor labour income estimates / "taxable income" estimates) from the PNAD surveys to our taxable income estimates from tax data for years when the latter comprised of pre-tax labour incomes, including pensions, and rent (1998–2016). When the taxable income concept includes more types of capital income (1974–1988), we use estimates from Morgan (2015) on the composition of top incomes to reduce them to pre-tax labour income and rent. Then we multiply them by the ratios from the surveys to arrive at estimates of factor income.[24]

Figure B.16 shows the gender gap by fractile of labour income in surveys and tax data. For the total employed population we can only use surveys, given that the tax data only covers a little more than 10% of the adult population. For top fractiles we can directly compare the gender gaps in both sources. As stated in the paper, women at the top of the distribution in Brazil have always had more comparable average earnings to men than women on average in the earning population. This finding is consistent whether we use tax data or survey data at least since 1998. In the main paper we extrapolate tax data estimates to years with missing data since 1976, using the average difference in the gender gap between survey estimates and tax data estimates for the years 1998, 2014 and 2015.

Figure B.17 shows the comparison of estimates for the share of women within top fractiles across data sources and their associated income concepts. Our desired income concept is pre-tax factor labour income for the income earning population, aged 20–64. Moreover, we tax data to be the most trusted data source. However, as explained, estimates from tax statistics are only available for the concept of taxable income for working and non-working adults (given that they include pension income). Thus, to convert these estimates to the concept we show in the paper we make use of survey data to compute

[24]The share of labour income and rent in total taxable income over 1974–1988 was about 92% on average for both the Top 10% and the Top 1%, increasing from 87–89% in 1974 to 94–95% by 1988 (see Morgan (2015)).

Figure B.16: Gender Gaps in Brazil: Surveys vs Tax Data



Notes: The gender gap is the average labour income of men divided by the average labour income of women. Distribution of factor labour income (wages/salaries, labour component of mixed income). Fractiles are defined relative to the total number of working-age individuals in the population (between 20 and 64 years of age). Authors' calculations from tax and survey data.

estimates that match the tax data concept (taxable income 20+) and estimates of our desired concept (factor labour income, 20–64). These are presented in Figure B.17 for women in the Top 10% and the Top 1%. Using these survey series we adjust the tax data series as explained at the beginning of this section. We effectively assume that the difference between the taxable income series and the factor labour income series in the surveys can be carried over to the tax data. This seems to us to be a reasonable approximation given the trends we observed. Unsurprisingly, the shares of women in top taxable incomes are generally higher than for pure labour income, largely due to the effect of pensions. We also show estimates for the shares of women at the top for the concept of total individual income among working adults from the decennial censuses. While levels are generally lower, given the inclusion of property incomes, the trends from the other series are replicated. These findings make our conclusions quite robust.

They also lighten the possibility that our results are driven by quirks specific to the tax data. These have to do with the nature and rules of filing a tax declaration in Brazil. In theory, two issues could be contaminating the validity of the results: the pooling of couple income and the division of their income in such a way that optimises their household's total tax liability.

Concerning the first, couples in Brazil have the *option* to return a joint declaration of their income to the tax authorities. This option allows them to benefit from a fixed modest deduction for the spouse as a dependent and also to pool their income for the application

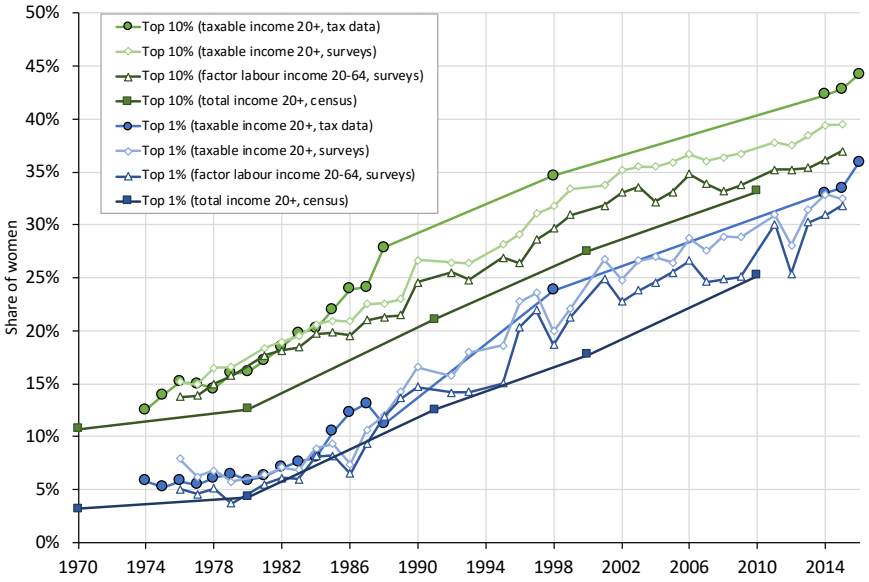
of the tax schedule if they are both earning (as opposed to being taxed separately). Thus, women filing a declaration could in theory have their spouses' income adjoined, similarly for the case of filing men. However, given the tax law, if a married woman has little income relative to her spouse, there are more incentives to file jointly if the allowance for dependants (including spouses with or without income) that is deductible from gross taxable income is greater than the additional tax due brought about by a joint declaration. If a woman has higher income then incentives increase for her to file separately, as her income would be subject to the different marginal tax rates (including the first exempt threshold) as opposed to being all subject to the highest rate, if she filed jointly with her partner. While more women would likely appear on the declaration of their husbands, rather than the reverse (given the higher average income of men), the nature of the tax incentives and the law for filing means that most high earning women that file a declaration are likely to be single or married and filing separately. In any case only 20-30% of all declarations are filed jointly, including for the highest income brackets, and male declarations in the Top 1% have between 2 and 3 times as many dependants on average than female declarations, suggesting that more women appear on male declarations than the opposite (as well as other relatives).

On the other hand, it could be possible that couples split their income so as to minimise their total tax liability. In practice, this would only be possible where couples work in the management of the same firm, so that they could organise their remuneration accordingly. This may be true for some high earners, but it is unlikely to be a very large share. In any case, since we are only concerned with factor labour incomes this type of practice would only affect salaries, and women would be reporting accurately what indeed they "earn". What is neither a real issue is the concern that some women appearing in higher income brackets only receive non-labour incomes, in line with their husbands allocating them investment income. This is because the gender-decomposed tax tabulations we work from rank income according to taxable income, which excludes non-labour incomes (apart from rent, which we adjust as explained above). Thus, we can conclude that our estimates are equally robust to a set of concerns based on potential artefacts of the tax data. Only if we get access to the micro-data can we re-evaluate these questions.

9.2 Factor Income Inequality

We compute shares for 2000–2015 using numerators and denominators from detailed sector level national accounts (IBGE, 2017). The difference between gross and net capital shares is the deduction of the consumption of fixed capital, which we estimate as described in section 5.2. Capital factor income includes the operating surplus of each economic sector as well as the capital component of mixed income. Labour factor income includes employee compensation of the household sector plus the labour component of mixed income. For 2000–2015, we use the 70–30 to attribute mixed income between labour and capital. This

Figure B.17: Representation of Women at the Top in Brazil: Comparison Across Data Sources and Concepts



Notes: Graph shows the share of employed women at the top of the labour income distribution. Employed population covers individuals between 20 and 64 years of age with positive factor labour income (wages/salaries, labour component of mixed income). Fractiles are defined relative to the total number of employed individuals in the population. Authors' calculations from tax, survey and census data.

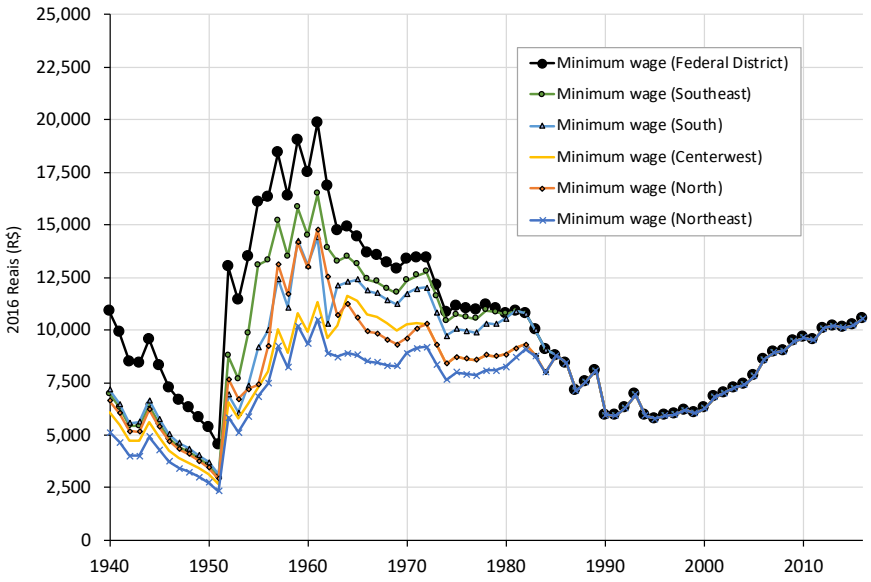
ratio corresponds to the average ratio of the labour remuneration and capital withdrawals of self-employed workers and micro-entrepreneurs calculated from the occupational tables in the tax statistics between 2007 and 2016 (Secretaria da Receita Federal, 2018). Labour remuneration is proxied by gross taxable income (mainly wages for this group of workers), while capital withdrawals are proxied by non-taxable income (mainly distributed profits, dividends and partner income for these workers).

To estimate factor shares before 2000, we anchor our 2000–2015 series to the annual growth of the labour share directly computed by Frankema (2010) for the period 1920–2000. The author computes total labour income as the sum of the average labour income per worker of the rural, urban-informal and urban-formal sectors multiplied by their respective labour force shares, using various sources (described in the Appendix of Frankema (2010)). This series effectively incorporates the wage component of mixed income using the "labour method", which attributes to the self-employed a wage equal to that of the average employee in each specific sector of economic activity (rural, urban informal, urban formal; in agriculture, industry and service-related activities).

9.3 Minimum Wages

Regional minimum wages for Brazil are presented in Figure B.18. The first minimum wages were introduced in urban areas at the state-district-level in 1940 for urban workers only. Each state comprised between 2 and 5 districts, depending on the population size. A minimum wage for rural workers was introduced in 1963. The Federal minimum wage was introduced in 1984, unifying previously distinct minimum levels.

Figure B.18: Minimum Wages by Region in Brazil



Notes: The federal minimum wage was introduced in 1984. Before 1984 minimum wages were defined by law at the state-district level, and only in urban areas before 1963. Authors calculations using nominal wage data from Saboia (1984). The Federal District comprises the city of Rio de Janeiro up to 1960 and Brasilia thereafter. All incomes are deflated by the GDP deflator from IBGE, and annually cumulated by multiplying monthly values by 12. See section 9.3 for more details.

For the earlier period we divide up the country into its five regions plus the Federal District, which comprised the capital city. Prior to 1961 the Federal District corresponds to the city of Rio de Janeiro, while from 1961 the Federal District changed to the new capital city, Brasilia. In our computations, northern regions comprise of the following states: Acre, Amazonas and Pará. Northeastern states are: Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe and Bahia. Southeastern states are: Minas Gerais, Espirito Santo, Rio de Janeiro and São Paulo. Southern states are: Paraná, Santa Caterina and Rio Grande do Sul. Finally, centerwest states are: Mato Grosso (including Mato Grosso do Sul) and Goiás. The minimum wage at the state level is the average of the district level minimum wages.

Nominal wage data at the state-district level are from Saboia (1984), who retrieves data from official sources. To calculate real minimum wages per region we take the ratio of each

state-level nominal minimum wage to the federal district minimum wage and multiply it by the federal minimum wage in 2016 prices (deflated by the GDP deflator). This implicitly assumes that the differences in the nominal values between states adequately captures differences in price levels, which seems a reasonable first approximation.

Before 1984 the Federal District had the highest state-level minimum wage observed in the country, alongside São Paulo city after 1964. The lowest state-level minimum wage is of the state of Maranhão in the Northeast.

Bibliography

Aaberge, R. and A. B. Atkinson (2010). Top incomes in norway. *Top incomes: a global perspective 2*.

Abreu, M. P. (1990). Crise, crescimento e modernização autoritária: 1930–1945. In M. de Paiva Abreu (Ed.), *A ordem do progresso: cem anos de política econômica republicana, 1889–1989*. Rio de Janeiro: Editora Campus.

Alstadsæter, A., M. Jacob, W. Kopczuk, and K. Telle (2016). Accounting for business income in measuring top income shares: Integrated accrual approach using individual and firm data from norway. *NBER Working Paper* (22888).

Alston, L., M. A. Melo, B. Mueller, and C. Pereira (2013). Changing social contracts: beliefs and dissipative inclusion in Brazil. *Journal of Comparative Economics 41*, 48–65.

Alvaredo, F. (2010). The rich in argentina over the twentieth century, 1932-2004. *Top incomes: A Global Perspective 2*.

Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters 110*(3), 274–277.

Alvaredo, F., L. Assouad, and T. Piketty (2017). Measuring Inequality in the Middle East 1990-2016: The World's Most Unequal Region? (September).

Alvaredo, F., A. B. Atkinson, L. Chancel, T. Piketty, E. Saez, and G. Zucman (2017). Distributional National Accounts (DINA) Guidelines: Concepts and Methods used in WID.world. *WID.world Working Paper Series 2017/6* (June).

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018). *World inequality report 2018*. Belknap Press of Harvard University Press.

Amadeo, E., I. Gill, and M. Neri (2000). *Brazil: The Pressure Points in Labor Legislation*. Ensaios Econômicos, 395. Rio de Janeiro: Fundação Getúlio Vargas.

Atkinson, A. B. (2007). The distribution of top incomes in the united kingdom 1908–2000. In A. B. Atkinson and T. Piketty (Eds.), *Top Incomes over the Twentieth Century: A*

Contrast between Continental European and English-Speaking Countries, Volume 1, pp. 82–140. Oxford University Press.

Atkinson, A. B. and T. Piketty (Eds.) (2007). *Top incomes over the twentieth century: a contrast between continental European and English-speaking countries*. Oxford University Press.

Atkinson, A. B. and T. Piketty (2010). *Top incomes: A Global Perspective*. Oxford University Press.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Perspectives 49*(1).

Atria, J., I. Flores, R. Mayer, and C. Sanhueza (2018). Top incomes in chile. a historical perspective of income inequality (1964-2015).

Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman (1955). An Empirical Distribution Function for Sampling with Incomplete Information. *Ann. Math. Statist. 26*(4), 641–647.

Bacha, E. L. (1980). Selected issues in post-1964 brazilian economic growth. In L. Taylor, E. L. Bacha, E. A. Cardoso, and F. J. Lysy (Eds.), *Models of Growth and Distribution for Brazil*. Oxford University Press.

Baer, W. (2014). *The Brazilian Economy: Growth and Development*. London: Reinner.

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics 104*, 184–198.

Barros, R., M. De Carvalho, S. Franco, and R. Mendonça (2010). Markets, the state, and the dynamics of inequality in brazil. In L. F. López-Calva and N. C. Lustig (Eds.), *Declining Inequality in Latin America: A Decade of Progress?*, pp. 134–74. Brookings Institution Press.

Bengtsson, E. (2017). The swedish sonderweg in question: Democratization and inequality in comparative perspective, c. 1750–1920.

Bengtsson, E. and D. Waldenström (2018). Capital shares and income inequality: Evidence from the long run. *The Journal of Economic History 78*(3), 712–743.

Blanchet, T., J. Fournier, and T. Piketty (2017). Generalized Pareto Curves: Theory and Applications. *WID.world Working Paper 2017/3*(March).

Blankenburg, S. and J. G. Palma (2012). Economic development. In J. King (Ed.), *The Elgar Companion To Post-Keynesian Economics*, pp. 138–143. Edward Elgar.

Blyth, M. (2002). *Great transformations: Economic Ideas and Institutional Change in the Twentieth Century*. Cambridge University Press.

Bourguignon, F. (2018). Simple adjustments of observed distributions for missing income and missing people. *The Journal of Economic Inequality*, 1–18.

Brasil (1974). *II Plano Nacional de Desenvolvimento (1975–1979)*. Brasília: Imprensa Oficial.

Brito, A. S., M. Foguel, and C. L. Kerstenetzky (2016). Minimum wage valorization policy and the decline in income inequality in brazil between 1995 and 2014: A decomposition approach. *Texto para Discussão* (115).

Brunk, H. D. (1955). Maximum Likelihood Estimates of Monotone Parameters. *Ann. Math. Statist. 26*(4), 607–616.

Bruno, M., H. Diawara, E. Araújo, A. C. Reis, and M. Rubens (2011). Finance-led growth regime no brasil: estatuto teórico, evidências empíricas e consequências macroeconômicas. *Revista de Economia Política 31*(5), 730–750.

Bruns, B., D. Evans, and J. Luque (2011). *Achieving World-Class Education in Brazil: The Next Agenda*. The World Bank.

Burkhauser, R. V., M. H. Hahn, and R. Wilkins (2016). Top Incomes and Inequality in Australia: Reconciling Recent Estimates from Household Survey and Tax Return Data. *Melbourne Institute Working Paper Series*.

Burkhauser, R. V., N. Hérault, S. P. Jenkins, and R. Wilkins (2016). What has Been Happening to UK Income Inequality Since the Mid-1990s? Answers from Reconciled and Combined Household Survey and Tax Return Data. *NBER Working Paper* (21991).

Burkhauser, R. V., N. Hérault, S. P. Jenkins, and R. Wilkins (2018). Survey Under-Coverage of Top Incomes and Estimation of Inequality: What is the Role of the UK's SPI Adjustment? *Fiscal Studies 39*(2), 213–240.

Canberra Group (2011). *Canberra Group Handbook on Household Income Statistics* (2^{nd} ed.). Genebra: United Nations.

Carvalho, L. (2018). *Valsa brasileira: Do boom ao caos econômico*. Editora Todavia SA.

Castro, F. A. and M. Bugarin (2017). A progressividade do imposto de renda da pessoa física no brasil. *Estudos Econômicos 47*(2), 259–293.

Centro de Informações Econômico–Fiscais (1968). *Imposto sobre a renda e proventos de qualquer natureza 1968*. Rio de Janeiro: Ministério da Fazenda.

Chancel, L. and T. Piketty (2017). Indian income inequality, 1922-2014: From British Raj to Billionaire Raj? *WID.world Working Paper Series 2017/11*(September).

Costa, S. A. (1981). *O CGT e as lutas sindicais brasileiras (1960-64)*. Editora do Grêmio Politécnico.

Couri, D. V. and P. R. S. Bijos (2016). Breve análise sobre a pec 241/2016, que altera o ato das disposições constitucionais transitórias para instituir o novo regime fiscal.

Czajka, L. (2017). Income Inequality in Côte d'Ivoire: 1985-2014. *WID.world Working Paper 2017/8*(July).

Dean, W. (1971). Latifundia and land policy in nineteenth-century brazil. *The Hispanic American Historical Review 51*(4), 606–625.

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics 87*(1), 1–19.

Deming, W. E. and F. F. Stephan (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics 11*(4), 427–444.

Departamento do Imposto de Renda (1965). *Relatório das atividades do ano de 1964*. Brasília: Departamento de Imprensa Nacional.

Departamento do Imposto de Renda (1968a). *Boletim Estatístico, n. 28*. Ministério da Fazenda.

Departamento do Imposto de Renda (1968b, August). *Boletim Estatístico, n. 32*. Ministério da Fazenda.

Deville, J.-C. (2000). Generalized calibration and application to weighting for non-response. In J. G. Bethlehem and P. G. M. van der Heijden (Eds.), *COMPSTAT: Proceedings in Computational Statistics 14th Symposium held in Utrecht, The Netherlands, 2000*, pp. 65–76. Heidelberg: Physica-Verlag HD.

Deville, J.-C. and C.-E. Särndal (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association 87*(418), 376–382.

Divisão do Imposto de Renda (1966, October). *Boletim Estatístico, n. 23*. Ministério da Fazenda.

DWP (2015). Households Below Average Income: An analysis of the income distribution 1994/95 – 2013/14. (June).

Engerman, S. L. and K. L. Sokoloff (2012). *Economic Development in the Americas since 1500: Endowments and Institutions*. Cambridge University Press.

Fairfield, T. and M. Jorratt De Luis (2016). Top income shares, business profits, and effective tax rates in contemporary chile. *Review of Income and Wealth 62*, S120–S144.

Fernandes, R. C. (2016). Sistema tributário e desigualdade: uma análise do impacto distributivo do imposto de renda no brasil. Master's thesis in economics, Universidade Federal de Minas Gerais, Belo Horizonte.

Ferreira, F., P. Lanjouw, and M. Neri (2003). A robust poverty profile for brazil using multiple data sources. *Revista Brasileira de Economia 57*(1), 59–92.

Ferreira, F. H., S. P. Firpo, and J. Messina (2017). *Ageing Poorly? Accounting for the decline in earnings inequality in Brazil, 1995–2012*. The World Bank.

FGV (1962). *Contas Nacionais do Brasil, 1947/1961*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística.

Finlay, M. (1980). *Ancient Slavery and Modern Ideology*. New York: The Viking Press.

Fishlow, A. (1972, março). Brazilian size distribution of income. *The American Economic Review 62*(1/2), 391–402.

Fleming, K. G. (2007). We're Skewed—The Bias in Small Samples from Skewed Distributions. *Casualty Actuarial Society Forum 2*(2), 179–183.

Frankema, E. (2010). Reconstructing labor income shares in argentina, brazil and mexico, 1870-2000. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History 28*(2), 343–374.

Furtado, C. (1965). Political obstacles to economic growth in brazil. *International Affairs 41*(2), 252–266.

Furtado, C. (1969). Desenvolvimento e estagnação na américa latina: Um enfoque estruturalista. *América Latina: Ensayos de Dnterpretación Económica. Santiago: Ed. Universitária*, 120–149.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2016). Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA). (April).

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2018). Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA). *Journal of Public Economics 162*, 63–77.

Gethin, A. and M. Morgan (2018). Brazil divided: Hindsights on the growing politicisation of inequality. *WID.world Issue Brief 2018/3*(October).

Gobetti, S. and R. Orair (2015). Política fiscal em perspectiva: o ciclo de 16 anos (1999-2014). *Revista de Economia Contemporânea 19*(3), 417–447.

Gobetti, S. and R. Orair (2017a). Taxation and distribution of income in brazil: new evidence from personal income tax data. *Revista de Economia Política 37*(2), 267–286.

Gobetti, S. W. and R. O. Orair (2017b). Taxation and distribution of income in brazil: new evidence from personal income tax data. *Brazilian Journal of Political Economy 37*(2), 267–286.

Gomes, A. C. (2005). *A invenção do trabalhismo* (3^a ed. ed.). Rio de Janeiro: Editora FGV.

Gonzalez, R. H. (2010). *Employment policies in Brazil: history, scope and limitations.* IPC-IG Working Paper n. 70. Brasília: International Policy Centre for Inclusive Growth (IPC-IG).

Hallak Neto, J. a. and C. M. Forte (2016). O sistema de contas nacionais: evolução histórica e implantação no brasil. *Revista Econômica 18*(1).

Hirschman, A. O. and M. Rothschild (1973). The changing tolerance for income inequality in the course of economic development: With a mathematical appendix. *The Quarterly Journal of Economics 87*(4), 544–566.

Hlasny, V. and P. Verme (2017). The impact of top incomes biases on the measurement of inequality in the United States.

Hlasny, V. and P. Verme (2018). Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data Vladimir. *Econometrics 6*(30), 1–38.

Hoerl, A. E. and R. W. Kennard (2000). Ridge Regression: Biased Estimation for Problems Nonorthogonal. *Technometrics 42*(1), 80–86.

Hoffmann, R. (1988). A subdeclaração dos rendimentos. *Revista São Paulo em Perspectiva 2*(1), 50–54.

Hoffmann, R. and J. C. Duarte (1972). A distribuição da renda no brasil. *Revista de administração de empresas 12*(2), 46–66.

Hoffmann, R. and M. G. Ney (2010). Estrutura fundiária e propriedade agrícola no brasil, grandes regiões e unidades da federação. *Brasília: Ministério do Desenvolvimento Agrário*, 108.

IBGE (1952). *Censo Demográfico 1950.* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística.

IBGE (2000). *Sistema de Contas Nacionais - Brasil, 1995-1999.* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística.

IBGE (2006). *Estatísticas do século XX.* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística.

IBGE (2017). *Anuário Estatístico do Brasil.* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística. (72 volumes; yearly since 1945).

IBGE (2017). *Sistema de Contas Nacionais: Brasil: 2015.* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística.

Jaramillo, M. and J. Saavedra (2005). Severance payment programs in latin america. *Empirica 32*, 275–307.

Jenkins, S. P. (2017). Pareto Models, Top Incomes and Recent Trends in UK Income Inequality. *Economica 84*(334), 261–289.

Kaldor, N. (1955). Alternative theories of distribution. *The Review of Economic Studies 23*(2), 83–100.

Kaldor, N. (1957). Inflação e desenvolvimento econômico. *Revista Brasileira de Economia 11*(1), 55–82.

Kane, C. T. and J. Morisset (1993). *Who would vote for inflation in Brazil?: an integrated framework approach to inflation and income distribution*, Volume 1183. World Bank Publications.

Kennickell, A. B. (2009). Getting to the top: reaching wealthy respondents in the SCF. *Proceedings of the Section on Survey Research Methods, American Statistical Association Joint Statistical Meetings n/a*, 2055–2069. Acesso 11 abr. 2016.

Keynes, J. M. ([1919] 2011). *The Economic Consequences of the Peace.* Digireads.com.

Keynes, J. M. ([1936] 1964). *The General Theory of Employment, Interest and Money.* London: Harcourt Brace.

Kingston, J. and L. Kingston (1972). A distribuição da renda no brasil, 1960-1970. *Revista Brasileira de Economia 26*(4), 241–256.

Korinek, A., J. A. Mistiaen, and M. Ravallion (2006). Survey nonresponse and the distribution of income. *Journal of Economic Inequality 4*(1), 33–55.

Kuznets, S. (1953). *Shares of Upper Income Groups in Income and Savings*. NBER.

Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review 45*(1), 1–28.

Lakner, C. and B. Milanovic (2013). *Global income distribution: from the fall of the Berlin Wall to the Great Recession*. The World Bank.

Langoni, C. (1973). *Distribuição da renda e desenvolvimento economico do Brasil*. Rio de Janeiro: Editora Expressao e Cultura.

Lazonick, W. (2014). Profits without prosperity. *Harvard Business Review 92*(9), 46–55.

Lesage, E., D. Haziza, and X. D'Haultfoeuille (2018). A cautionary tale on instrument vector calibration for the treatment of unit nonresponse in surveys.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The manchester school 22*(2), 139–191.

López-Calva, L. F. and N. C. Lustig (2010). *Declining Inequality in Latin America: A Decade of Progress?* Brookings Institution Press.

Love, J. L. (1970). Political participation in brazil, 1881-1969. *Luso-Brazilian Review 7*(2), 3–24.

Lustig, N., L. F. Lopez-Calva, E. Ortiz-Juarez, and C. Monga (2016). Deconstructing the decline in inequality in latin america. In *Inequality and growth: Patterns and policy*, pp. 212–247. Springer.

Martine, G. (2005). O papel dos organismos internacionais na evolução dos estudos populacionais no brasil: notas preliminares. *Revista Brasileira de Estudos de População 22*(2), 257–275.

Martine, G., A. A. Camarano, P. Fletcher, and R. Neupert (1988). A pnad: Notas para uma avaliação. In D. Sawyer (Ed.), *PNADs em foco: anos 80*, pp. 281–305. Belo Horizonte: Abep.

Marx, K. ([1887] 1990). *Capital. Volume 1: A Critique of Political Economy*. Penguin Classics.

Medeiros, M., J. de Castro Galvão, and L. de Azevedo Nazareno (2018). Correcting the underestimation of top incomes: Combining data from income tax reports and the brazilian 2010 census. *Social Indicators Research 135*(1), 233–244.

Medeiros, M., P. Souza, and F. Castro (2015a). O topo da distribuição de renda no brasil: Primeiras estimativas com dados tributários e comparação com pesquisas domiciliares (2006-2012). *Revista de Ciências Sociais 58*(1).

Medeiros, M., P. H. G. F. Souza, and F. A. d. Castro (2015b). The stability of income inequality in Brazil, 2006–2012: an estimate using income tax data and household surveys. *Ciência & Saúde Coletiva 20*(4), 971–986.

Médici, A. (1988). A mensuração da subjetividade: notas sobre a variável renda nas PNADs. In D. Sawyer (Ed.), *PNADs em foco: anos 80*. Belo Horizonte: Abep.

Morais, M. d. P. and B. Cruz (2003). *Demand for housing and urban services in Brazil: a hedonic approach*. Texto para Discussão n. 946. Instituto de Pesquisa Econômica Aplicada.

Morandi, L. and E. Reis (2004). Estoque de capital fixo no brasil, 1950-2002. *Anais do XXXII Encontro Nacional de Economia 42*.

Morgan, M. (2015). Income concentration in a context of late development: an investigation of top incomes in brazil using tax records, 1933-2013. Master's dissertation, Paris School of Economics, Paris.

Morgan, M. (2017). Extreme and Persistent Inequality: New Evidence for Brazil Combining National Accounts, Surveys and Fiscal Data, 2001-2015. *WID.world Working Paper 2017/12*.

Morgan, M. (2018). The expansion of social spending in brazil 1932-2015: The emergence of a new welfare-state growth paradigm?

Morley, S. A. (1978). Growth and inequality in brazil. *Luso-Brazilian Review 15*(2), 244–271.

Mortara, G. (1949a). Dados e observações sôbre a distribuição das rendas das pessoas físicas contribuintes do impôsto de renda no Brasil. *Revista Brasileira de Estatística 10*(38), 197–222.

Mortara, G. (1949b). Representação analítica das distribuições dos contribuintes e das respectivas rendas líquidas determinadas para a aplicação do impôsto de renda, em função do valor da renda líquida. *Revista Brasileira de Economia 3*(2), 7–34.

Nóbrega, C. B. (2014). *História do Imposto de renda no Brasil, um enfoque da pessoa física (1922–2013)*. Brasília: Receita Federal.

Nunes, E. P. (1988). *Sistema de Contas Nacionais: A Gênese das Contas Nacionais Modernas e a Evolução das Contas Nacionais no Brasil*. Phd dissertation in economics, Universidade de Campinas, Campinas.

Okolewski, A. and T. Rychlik (2001). Sharp distribution-free bounds on the bias in estimating quantiles via order statistics. *Statistics and Probability Letters 52*(2), 207–213.

Palma, J. G. (2011). Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-u': It's all about the share of the rich. *Development and Change 42*(1), 87–153.

Palma, J. G. (2016). Do nations just get the inequality they deserve? the "palma ratio" re-examined. In K. Basu and J. E. Stiglitz (Eds.), *Inequality and Growth: Patterns and Policy*, pp. 35–97. Springer.

Paoli, M. C. (1989). Trabalhadores e cidadania: experiência do mundo público na história do Brasil moderno. *Estudos Avançados 3*(7), 40–66.

Papke, L. and J. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics 11*, 619–632.

Papke, L. and J. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*, 121–133.

Pareto, V. (1896). *Écrits sur la courbe de la répartition de la richesse.*

Piketty, T. (2003). Income Inequality in France, 1901–1998. *Journal of Political Economy 111*(5), 1004–1042.

Piketty, T. and E. Saez (2003). Income inequality in the United States, 1913-1998. *Quarterly Journal of Economics 118*(1), 1–39.

Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American economic journal: economic policy 6*(1), 230–71.

Piketty, T., E. Saez, and G. Zucman (2018). Distributional National Accounts: Methods and Estimates for the United States*. *The Quarterly Journal of Economics 133*(2), 553–609.

Piketty, T., L. Yang, and G. Zucman (2017). Capital Accumulation, Private Property and Rising Inequality in China, 1978-2015. *WID.world Working Paper Series 2017/6*(April).

Piketty, T. and G. Zucman (2014). Capital is Back: Wealth-Income Rations in Rich Countries 1700–2010. *Quarterly Journal of Economics 129*(3), 1255–1310.

Rocha, S. (2003). A investigação da renda nas pesquisas domiciliares. *Economia e Sociedade 12*(2), 205–224.

Saad-Filho, A. and M. d. L. R. Mollo (2002). Inflation and stabilization in brazil: A political economy analysis. *Review of Radical Political Economics 34*(2), 109–135.

Saboia, J. a. (1984). *Evolução histórica do salário mínimo no Brasil: fixação, valor real e diferenciação regional.* Rio de Janeiro: IPEA, Programa Nacional de Pesquisa Econômica.

Scheidel2017 (2017). *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*. Princeton, NJ: Princeton University Press.

Secretaria da Receita Federal (1970–1980). *Anuário Econômico–Fiscal, anos 1–11*. Rio de Janeiro; Brasília: Ministério da Fazenda. (Annual, 1970–1987).

Secretaria da Receita Federal (1989). *Imposto de Renda – Pessoa Física, anos 1–11*. Brasília: Ministério da Fazenda. (Annual, 1979-1989).

Secretaria da Receita Federal (2000). *Análise econômica da DIRPF 1999*. Texto para Discussão n. 9. Brasília: Ministério da Fazenda.

Secretaria da Receita Federal (2001). *Tributação da renda no Brasil Pós–Real*. Brasília: Secretaria da Receita Federal.

Secretaria da Receita Federal (2004). *O imposto de rendas das pessoas físicas no Brasil*. Estudos Tributários n. 14. Brasília: Secretaria da Receita Federal.

Secretaria da Receita Federal (2018). Grandes Números IRPF – anos-calendário 1999 a 2016. Acessed 9 Mar. 2018.

Serra, J. and M. d. C. Tavares (2000). Além da estagnação: Uma discussão sobre o estilo de desenvolvimento recente do brasil. *En: Cinqüenta anos de pensamento na CEPAL-Rio de Janeiro: Record/CEPAL, 2000-v. 2, p. 589-608*.

Silva, J. M. P. Q., C. V. d. Carvalho, et al. (2015, November). Tributação de lucros e dividendos no brasil: uma perspectiva comparada.

Simonsen, M. H. (1975). A política salarial dos governos da revolução. *Brasília: House of Representatives Committee on Wage Policy, November 18; processed*.

Singh, A. C. and C. A. Mohl (1996). Understanding Calibration Estimators in Survey Sampling. *Survey Methodology 22*(2), 107–115.

Skidmore, T. E. (1976). *Brasil: de Get/úlio Vargas a Castelo Branco (1930–1964)*. Rio de Janeiro: Paz e Terra.

Soares, S. (2017). *Valor de mercado de acesso a eletricidade, água, esgoto, gás e telefonia fixa*. Texto para Discussão n. 2317. Instituto de Pesquisa Econômica Aplicada.

Soares, S. and A. Lima (2002). *A mensuração da educação nas PNADs da década de 1990*. Texto para Discussão n. 928. Brasília, Brazil: Ipea.

Soares, S. S. D. (2008). *O ritmo de queda na desigualdade no Brasil é adequado? Evidências do contexto histórico e internacional*. Texto para Discussão n. 1339. Brasília: Instituto de Pesquisa Econômica Aplicada.

Soares, S. S. D., P. H. G. F. Souza, R. G. Osorio, and F. G. Silveira (2010). Os impactos do benefício do Programa Bolsa Família sobre a desigualdade e a pobreza. In J. A. de Castro and L. Modesto (Eds.), *Bolsa Família 2003–2010: avanços e desafios*, Volume 2. Brasília: Instituto de Pesquisa Econômica Aplicada.

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What Are We Weighting For? *Journal of Human Resources 50*(2), 301–316.

Souza, P. H. G. F. (2013). Uma metodologia para explicar diferenças entre dados administrativos e pesquisas amostrais, com aplicação para o bolsa famĺia e o benefício de prestação continuada na pnad. *Revista Brasileira de Estudos de População 30*(1), 299–315.

Souza, P. H. G. F. (2015). A distribuição de renda nas pesquisas domiciliares brasileiras: harmonização e comparação entre Censos, PNADs e POFs. *Revista Brasileira de Estudos de População 32*(1), 165–188.

Souza, P. H. G. F. (2016). *A desigualdade vista do topo: a concentração de renda entre os ricos no Brasil, 1926–2013*. Phd dissertation in sociology, Universidade de Brasília, Brasília.

Souza, P. H. G. F. (2018). A history of inequality: top incomes in brazil, 1926–2015. *Research in Social Stratification and Mobility 57*, 35–45.

Souza, P. H. G. F. and M. Medeiros (2015). Top income shares and inequality in Brazil, 1928–2012. *Sociologies in Dialogue 1*(1), 119–132.

Souza Reis, F. T. (1930). *O imposto de renda em seis annos de adaptação no Brasil, 1924–1929*. Rio de Janeiro.

Taleb, N. N. and R. Douady (2015). On the super-additivity and estimation biases of quantile contributions. *Physica A: Statistical Mechanics and its Applications 429*, 252–260.

Taylor, L., E. L. Bacha, E. A. Cardoso, and F. J. Lysy (1980). *Models of Growth and Distribution for Brazil*. The World Bank.

Torres, A. d. M. (2003). Modelos de tributação ótima da renda: o cálculo de alíquotas marginais assintóticas para o Brasil no ano de 2000. Mestrado em economia do setor público, Universidade de Brasília, Brasília.

UN (2009). *System of National Accounts 2008*. New York: United Nations, European Commission, International Monetary Fund, Organization for Economic Cooperation and Development and World Bank.

van Eeden, C. (1958). *Testing and Estimating Ordered Parameters of Probability Distributions*. Ph. D. thesis, University of Amsterdam.

Vargas, G. (1951). *A Nova Política do Brasil. Vol. X*. Rio de Janeiro: José Olímpio.

Weinberg, D. H., C. T. Nelson, M. I. Roemer, and E. J. Welniak Jr (1999). Fifty years of U.S. income data from the Current Population Survey: alternatives, trends, and quality. *The American Economic Review 89*(2), 18–22.

Weisbrot, M., J. Johnston, J. V. Carrillo, et al. (2016). Brazil's enormous interest rate tax: Can brazilians afford it? Technical report, Center for Economic and Policy Research (CEPR).

Weyland, K. (1996). *Democracy without equity: Failures of reform in Brazil*. University of Pittsburgh Press.

Williamson, J. G. (2011). *Trade and poverty: When the Third World Fell Behind*. MIT press.

Williamson, J. G. (2015). Latin american inequality: Colonial origins, commodity booms or a missed twentieth-century leveling? *Journal of Human Development and Capabilities 16*(3), 324–341.