Top incomes' impacts on inequality, growth, and social welfare

Combining surveys and income tax data in Brazil

Marcelo Neri[1] and Marcos Hecksher[2]

October 2018

Abstract: This paper evaluates the impacts of combining household surveys with income tax return files, in terms of growth, inequality, and social welfare in Brazil from 2007 to 2015. This exercise holds the promise of adding more realistic top income values to traditional surveys. While the previous literature focused on the impacts of these data combination exercises on income inequality, we assess their cumulative welfare implications. First, as the level of inequality rises when higher top incomes replace previous estimates from surveys, this exercise also increases by construction the mean and social welfare levels. Second, while the movement of these combined estimates presents a slower inequality fall than pure surveys, mean and social welfare growth is seen to have been faster. We are able to reconcile most of the aggregate discrepancies between income tax returns, surveys, and GDP growth rates but demographic inconsistencies still remains. Finally, the paper analyses the nature and causes of a series of measurement issues—in particular, why exempt incomes drove the growth in income tax returns during this period.

Keywords: Top incomes income inequality, personal income tax records, combining data sets, Pareto interpolation
JEL classification: I31, H2, H22

Acknowledgements: We thank participants at the conference 'Income Redistribution and The Role of Tax–Benefit Systems in Latin America', held in Quito in July 2018, for their comments. This paper will be presented at the National Meeting of Brazilian Economists (Associação Nacional dos Centros de Pós-Graduação em Economia, ANPEC) in Rio de Janeiro, 2018. We also acknowledge the financial support provided by Rede de Pesquisa e Conhecimento Aplicado (RPCAP) from the Getulio Vargas Foundation (FGV).

[1] Getulio Vargas Foundation (FGV) – FGV Social and FGV EPGE (*Escola de Pós-Graduação em Economia*), Rio de Janeiro, Brazil, corresponding author: Marcelo.Neri@fgv.br; [2] Instituto de Pesquisa Econômica Aplicada (IPEA) Brasília, Brazil, and Escola Nacional de Ciências Estatísticas (ENCE), Rio de Janeiro, Brazil.

This study has been prepared within the UNU-WIDER project on 'Inequality in the Giants'.

1 Introduction

Recently, there have been a few studies in Brazil combining data from household surveys with administrative records of personal income taxes (PIT, *Imposto de Renda da Pessoa Física*), held by the Inland Revenue Service (Secretaria de Receita Federal), whose main benefit was to measure more accurately the income of the wealthiest in the population. These studies have applied to the Brazilian case methods that are increasingly widely used by authors such as Piketty (2014) and Atkinson (2015) to estimate income inequality series over the years and even centuries in different countries. Medeiros et al. (2015a, 2015b) have been pioneers in harmonizing these different sources of information in Brazil. More specifically, they have reconciled microdata from the National Household Sample Survey (PNAD) and tabulations exclusive to the income tax return files from 2006 to 2012. Through this harmonizing process, the authors combine data close to the 90th percentile that roughly singles out the 10 per cent richest from the rest by replacing PNAD's adult population highest income by the highest income reported in the PIT. The hypothesis here is that the latter source has a greater capacity to identify the highest incomes. After all, it is not incentive-compatible to declare unrealistically high incomes if it will attract higher taxes, and usually the highest incomes are not well represented in the PNAD sample.

This literature emphasizes the impact of this change in measurement on income inequality, although it does not detail the joint effect in mean income and in social welfare related to the same data combination exercise. In the mixed database, the level of mean income and of social welfare would be unequivocally higher by construction, not only for the social welfare functions according to the usual hypotheses found in the economic literature, but also in terms of a more general Pareto efficiency criterion—that is, everyone is better off, or at least remains the same as before. That stands for any point in time when comparing the mixed PNAD-PIT distribution with the single PNAD distribution. In short, if Brazil followed the income distribution pattern of the mixed PNAD-PIT database, it would be unequivocally better off than the country portrayed in household surveys. To be clear, we refer to a country more unequal but more prosperous or the same for all segments of the population. A similar story seems to apply to comparisons in income distribution across time. In other words, if inequality presents a slower falling trend, both mean income and social welfare measures increase in the period analysed in these papers. We need to analyse the related changes in social welfare, both in levels and in changes across time. In any case, in economic evaluation of income distributions, one should not look just at their second moment without considering the first moment.

The present paper aims to evaluate the implications of combining household surveys with income tax return files, particularly in terms of social welfare and aspects such as mean income, extrapolating the impacts on income inequality in the recent Brazilian scenario. We suggest several extensions with respect to the previous literature. First, we include the period after 2012, when the signs of a major Brazilian recession began to appear. Second, we test different fittings in the distributions and evaluate their implications by using different components of social welfare, including mean income and inequality indexes. Finally, we attempt to address jointly the reasons for the static and dynamic behaviour in the two first moments of income distribution among different databases. We try to reconcile the discrepancies between income tax returns, surveys, and GDP growth rates. Another related question that has not been properly analysed is: why is the growth rate of mean income in the PIT mostly driven by exempt incomes? And also: has the reported income increase been overestimated?

The paper is thus organized as follows: in the second section, we briefly review the literature. In the third and main section of the paper, we present the main hypotheses supporting the combination of PNAD with PIT statistics and its results in terms of inequality, mean, and social welfare measures based on income, both in levels and in rates of change. In the fourth section, we delve into the possible causes of the high rate of growth in PIT income, which may be the least probed issue in the related literature and lies behind the social welfare trends observed. The fifth section analyses disaggregated income sources, in particular the increase of exempt incomes. The sixth section addresses demographic issues and economic incentives that determine the change of the number of PIT declarations by age groups—an illustrative example of PIT tables challenging established demographic facts. The sixth section presents the main conclusions of the paper.

2 Short literature review

Analyses on inequality in Brazil, traditionally based on household surveys, can be misleading if the survey's microdata have measurement errors associated with underestimation of upper incomes, which are the most important ones in determining the inequality level according to the usual measures, such as the Gini and the Theil-T indexes. Medeiros et al. (2015a, 2015b), who analysed tax returns between 2006 and 2012, concluded that inequality fell at a much slower rate than suggested by Brazilian household surveys in the same period. Their conclusion for the 2006–2012 period is reinforced by Souza (2016), who applied different methods to use tax data either in isolation or in conjunction with household data to create new Gini index series.

These works were pioneering in applying to Brazilian data—until then unavailable or unknown among the experts—some of the methods advocated by authors such as Piketty (2014) and Atkinson (2015), who used administrative records on PIT to estimate not only the level but also the variations in income inequality through the years and even over centuries. Integrating household survey information and income tax records is part of an important and promising research agenda that could reveal more accurately how income really is distributed within and among countries in the world.

The expectation that the PNAD particularly underestimates the highest incomes is justified by the patterns of differential non-response to the household survey. First, interviewers may have more obstacles in accessing the wealthiest households, which hinders their capacity to capture their income information. Second, the survey may have serious limitations in accurately measuring the wealthiest's share of income from sources such as rents, interests, and profits—as opposed to labour earnings or social benefits, which tend to have a pay cheque and are reported more accurately. Conversely, we have seen the possibility that in income tax return files, revenue from financial investments tends to be overestimated, as they do not consider the monetary correction, which is enclosed in the nominal variation of assets that banks report as income.

Although income tax return files may be the best tool to obtain an estimate closer to the real income level of the richest, they can be inadequate to estimate the variation rates of these incomes over time, which is crucial for the inequality trajectory. According to Medeiros et al. (2015a, 2015b), from 2006 until 2012, the income of the wealthiest population drifted away from the mean, contrary to what the PNAD indicates. This conclusion can only be legitimate if the underestimation of the wealthiest's income had increased in this period in the PNAD. This is perfectly possible, but we should not rule out other plausible explanations in view of the information disclosed so far.

The number of income tax return files corresponds to a fifth of the country's adult population. The wealthiest are overrepresented in the database, but it is evident that not every rich person is in the database; nor does each file contain the whole income of each taxpayer, and there may be overestimated incomes as well. In particular, not all incomes declared imply higher taxes paid, since there are tax-exempt income sources. Measurement error may also vary through the years. Incentives and the choice to file or not file each fraction of one's income may be influenced by legal factors, as in the case of well-paid workers who acquire the status of legally established firms, whose profits pay low business taxes and are also exempt from personal income taxes (Afonso 2017). Tax collection initiatives and enforcement also improved over this period. All these factors can increase the share of the wealthiest population's income captured by the Inland Revenue Service (IRS), thus influencing its growth rate.

A possible future disclosure of more detailed information may indicate a strong growth in capital income, which is clearly overlooked in the PNAD data. However, the path of the net operational surplus's share in the functional income distribution, according to the National Accounts, does not validate this hypothesis (Bastos 2012; Saboia and Hallak Neto 2014; Barros 2016). Nor does the path of the distribution of estimated real estate value based on household surveys (Neri 2014).

All of this discussion is important to put the forthcoming estimates into perspective. After all, it is not a simple task to reveal the 'real path' of income distribution in Brazil, which is the leitmotiv of these studies. Our aim is to quantify the level and changes in social welfare in Brazil based on the PNAD-PIT combined database. Then we study its immediate determinants, such as income inequality and mean income. The next step is to analyse details of income distribution gains and losses in the period according to each data source analysed, incorporating their immediate determinants.

3 Combining household surveys and income tax files

3.1 Combination

We have chosen to combine PNAD data with income tax files tabulations according to the methods first used in Brazil by Medeiros et al. (2015b). Between the first (2007) and the last (2015) PIT tabulation as disclosed by the IRS on the internet in a standard format[1], the wealthiest population's personal income was calculated and then combined with income in the PNAD for the remaining population. Afterwards, once we had combined the databases from the PNAD and the PIT, we analysed changes not only in inequality but also in the mean and in social welfare-related statistics.

The income tax tables disclosed on the IRS website present taxpayers' files in 2008 and 2016 regarding their income in 2007 and 2015, respectively. All incomes, deductions, taxes, assets, and liabilities are presented in each table for different categories, such as gender, city, age group, nature of labour, type of occupation, and monthly income ranges expressed in minimum wages (MW). The tables used in this section summarize total incomes from 'up to ½ MW' to 'more than 160 MWs' (in 2007) or 'more than 320 MWs' (in 2015) in monthly terms.[2]

---

[1] http://idg.receita.fazenda.gov.br/dados/receitadata/estudos-e-tributarios-e-aduaneiros/estudos-e-estatisticas/11-08-2014-grandes-numeros-dirpf/grandes-numeros-dirpf-capa (accessed 2017); various years available.

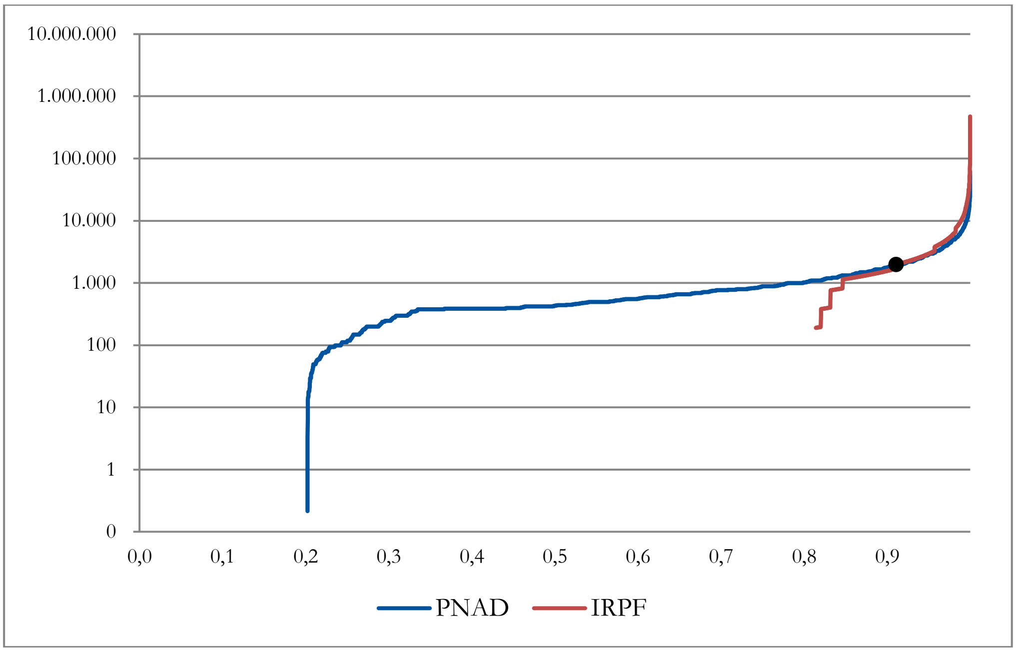
[2] The information disclosed has increased in detail in recent years, but more can be done if unidentified microdata integrating data on people and companies are also disclosed. Longitudinal microdata samples allow the same people to be tracked over the years, while respecting confidentiality, and different features of individual income processes to be estimated.

To use these aggregate figures to estimate a continuous distribution of income for each quantile of the population, a group of hypotheses of various degrees of realism is necessary. The main one is that the total income of each richest adult in the country, up to a certain percentage of the population, is correctly filed in the IRS database. The objective of this estimate is not to point out the 'real' value of income, but just to mitigate any underestimation in the household surveys.

According to Souza (2016), the total population is based on the Brazilian Geographical and Statistical Institute (IBGE) estimate for the number of people 20 years old or older living in Brazil on 1 July each year. In line with the above-mentioned authors and others, we applied a Pareto interpolation on these datasets to estimate the distribution within each income group.

Once the model had been chosen to estimate the highest income based on income tax files, the next step was to integrate them into estimates for the lowest income in the PNAD. There are countries where income tax files cover more than 90 per cent of the population. In Brazil, they correspond approximately to a fifth of the adult population—and not the richest fifth. Although they also provide fittings into alternative quantiles, Medeiros et al. (2015b) affirm that, based on data from different years, the income of the richest tenth can be estimated according to the income tax files, and the income of remaining nine-tenths according to the PNAD. Figure 1 shows excerpts where the PNAD and PIT estimates overlap in 2007. The chosen fitting points require an increase in the PNAD's income of the 8.9 per cent richest in 2007, while in 2015 alternative fitting points were adopted at the 11.4 per cent, 10.0 per cent, or 8.9 per cent richest to verify the robustness of the results found.

Figure 1: Individual monthly income by population quantile in 2007 (R$)



Source: Authors' construction from PNAD/IBGE; PIT/IRS.

Once these absolute fittings in the distributions have been applied[3], replacing the highest individual incomes declared in the PNAD by estimates based on the highest incomes as filed with the IRS increases the share of total income concentrated among the rich as well as the inequality among

[3] This is the fitting method adopted by Medeiros et al. (2015b). In order to include periods of high inflation in the analysis, Souza (2016) chose another method for estimating the inequality, based on the Gini index's capacity to split into groups that do not overlap.

the richest. These two changes increase the inequality in the combined PNAD-PIT database in relation to the PNAD.

Tables 1–3 show that the choice between the different fitting points considered in 2015 barely affects the results. While according to the PNAD, average real income increases by 1.7 per cent per year, the average real income of the PNAD-PIT database increases by 2.9 per cent per year considering the three fittings.

The Gini index, which falls at an average pace of 0.005 points per year in the PNAD, decreases by only 0.001 point per year in the PNAD-PIT database. The variation of the Theil-T index—more sensitive to the variations in the income of the richest than the Gini index—shifts according to the database: it drops by 2.7 per cent per year in the PNAD, but increases by 4.0 per cent per year in the PNAD-PIT database fitted in the 0.911 quantile. This last rate remains 3.9 per cent when the fitting is done in the 0.886 or 0.900 quantiles, which does not qualitatively alter the analysis.

Table 1: Mean income (constant R$ at 2015 prices)

	2007	2015	Total var.	Annual var.
PNAD	1,333	1,521	14.2%	1.7%
Fit 0.911	1,675	2,100	25.4%	2.9%
Fit 0.900		2,107	25.8%	2.9%
Fit 0.866		2,108	25.9%	2.9%

Source: Authors' construction from PNAD/IBGE; PIT/IRS and combined databases.

Table 2: Inequality (Gini)

	2007	2015	Total var.	Annual var.
PNAD	0.625	0.582	-0.043	-0.005
Fit 0.911	0.698	0.690	-0.008	-0.001
Fit 0.900		0.690	-0.008	-0.001
Fit 0.866		0.690	-0.008	-0.001

Source: Authors' construction from PNAD/IBGE; PIT/IRS and combined databases.

Table 3: Inequality (Theil-T)

	2007	2015	Total var.	Annual var.
PNAD	1.902	1.533	-19.4%	-2.7%
Fit 0.911	19.738	27.021	36.9%	4.0%
Fit 0.900		26.836	36.0%	3.9%
Fit 0.866		26.808	35.8%	3.9%

Source: Authors' construction from PNAD/IBGE; PIT/IRS and combined databases.

There was a strong increase in the incomes filed for taxes between 2007 and 2015. This growth in income tax was more intense than the average growth observed in the PNAD, in which the richest presented a growth lower than average and inequality fell. Replacing the income of the richest in the PNAD by figures provided by income tax files thus results in a higher annual growth rate for the PNAD-PIT database (2.9 per cent) than the PNAD average (1.7 per cent). However, even in the PNAD-PIT database, average growth remains lower than the growth in the median income (3.0 per cent), which is unaffected by the integration of the database.

Merely observing these three rates is not enough to understand what happened to inequality and social welfare between 2007 and 2011. If the richest earned more than the mean, which is usually associated with increases in inequality but also social welfare, the median income also grew more than the average, which is usually linked to reductions in inequality but once again increases in aggregate welfare. Therefore, what happened to both dimensions?

3.2 Social welfare

All the tables in this section derive from previous ones. For simplicity purposes, we only apply a connection link between the 0.911 bases because the remaining ones present identical substantive results, or similar to the other analysed links. Nevertheless, we focus on the details in shifts, not only through years for the same concept but also among concepts for the same year.

We refer specifically to the social welfare measure proposed by Amartya Sen (1974), which results from multiplying mean income by the Gini inequality index complement. The welfare level—which in 2007 was 1.02 per cent higher for the combined distribution vis-à-vis single PNAD—increases to 2.41 per cent, thus causing the welfare gain between 2007 and 2015 as presented by the combined database to grow by 3.2 per cent per year against 3.0 per cent per year in the PNAD. This superior performance of the country according to Sen's measure happens because—despite inequality having decreased 0.7 pp less in the combined database—the income growth was 1.2 pp higher each year. Therefore, by this criterion, looking at both shifts in the first two moments of income distribution yields higher welfare growth in the combined database than in the single PNAD.

Table 4: Social welfare Gini index based (Sen 1974)

	2007	2015	Total var.	Annual var.
PNAD	500	636	27.2%	3.0%
Fit 0.911	505	651	28.9%	3.2%
Across bases	1.02%	2.41%	1.7%	0.2%

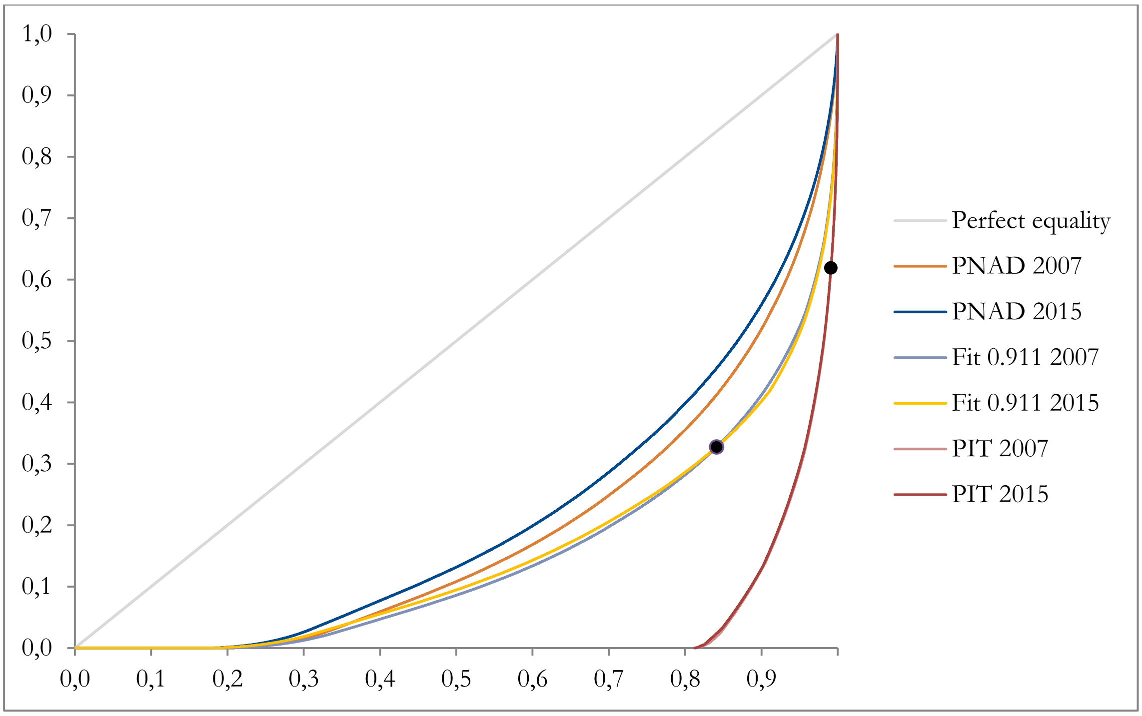
Source: Authors' construction from PNAD/IBGE and combined PNAD-PIT databases.

3.3 Inequality

As the PNAD-PIT combination increases the income of the richest in relation to the income observed in the PNAD without altering the income of the remaining population, the inequality level necessarily remains higher in the combined database, whatever the inequality index used. The degree of this increase in inequality as a result of the database integration thus depends on the inequality index. Moreover, variation in the index depends on its sensitivity to each part of the distribution. The Gini index is more sensitive to variations in quantiles closer to the median, while the Theil-T index is more sensitive to variations in the higher incomes. This allowed the PNAD-PIT inequality to decrease between 2007 and 2015 according to the Gini index, while it increased according to Theil-T.

The Lorenz curves (Figure 2) show, in the vertical axis, the percentage of income accumulated by the poorest population up to each respective quantile presented in the horizontal axis. If all people had the same income, the Lorenz curve would coincide with the straight line of perfect equality presented in the graph. The more distant the curve gets from the straight line, the greater the inequality in the distribution, but each index has a distinct sensitivity to different stretches of the curve. The PNAD 2007 curve is completely below the PNAD 2015 curve, which proves that inequality fell in this period, whatever the index used.

Figure 2: Lorenz curves for PNAD, PIT, and PNAD-PIT databases in 2007 and 2015



Source: Authors' construction from PNAD/IBGE; PIT/IRS and combined databases.

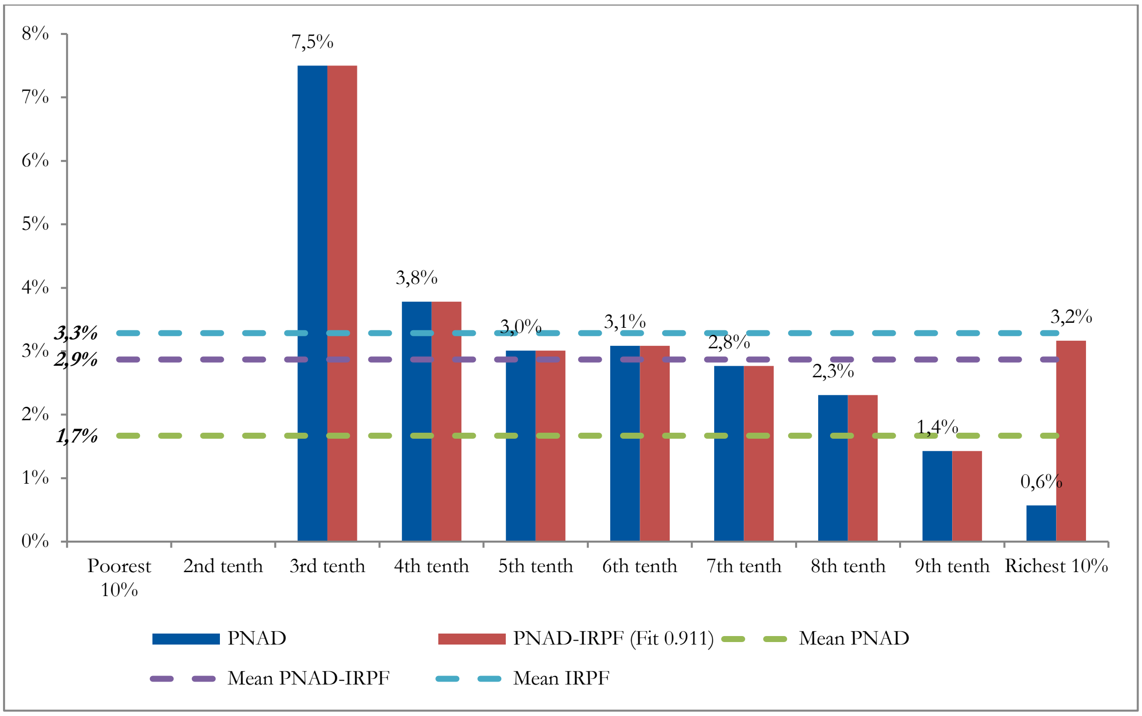
The PNAD-PIT integration produces Lorenz curves more distant from perfect equality, but the curves for the integrated bases in 2007 and 2015 meet each other. When two Lorenz curves intersect, there are indexes indicating a rise in inequality, while other indexes will indicate a decrease. This was exactly what happened, respectively, with the Theil-T and Gini indexes.

Figure 2 shows that the crossing occurs in the 0.842 quantile, where both curves have the same height of 0.328. That is, the 84.2 per cent poorest still had 32.8 per cent of the total income, while the 15.8 per cent richest retained 67.2 per cent of the total. Nonetheless, the 2007 curve is higher than 2015's in the whole of its extension to the left of the crossing point, meaning that the poorest increased their share in total income and experienced a growth higher than the average. Conversely, the 2015 curve is completely below 2007's in the whole of its extension to the right of the crossing point, meaning that the richest also earned more than the average and also increased their share in the total income.

3.4 Income distribution

The finding on inequality changes leads to the inevitable question: who then lost their share in the total income? In the PNAD-PIT integrated database, the seventh, eighth, and ninth tenths were the only ones that lost their shares in total income because their growth rates were lower than average (2.9 per cent per year), as shown in Figure 3.

Figure 3: Real growth rate of income per tenth of the population per year (2007–15)



Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

It is worth mentioning that the tenth with the lowest growth was the ninth (1.4 per cent), even though its growth was still positive. This group comprised people with income close to the average value, which corresponded to the 0.802 quantile in 2007 and the 0.817 quantile in 2015.

The two poorest tenths do not feature in the graph of growth rates because 20.2 per cent of adults had null income in 2007. This percentage dropped to 17.7 per cent in 2015. In other words, 2.5 per cent did not have income in 2007 and then earned income in 2015, but the graph does not display their earnings (infinite in percentage terms) nor the stagnation of the 17.7 per cent who remained with a null income between 2007 and 2015. The graph reveals that the poorest up to the 0.6 quantile of the adult population increased their share of total income. The 10 per cent richest also had a growth rate (3.2 per cent) higher than average (2.9 per cent) but not as high as the rates observed in the fourth and fifth tenths (7.5 per cent and 3.8 per cent, respectively).

The 10.1 goal in the Sustainable Development Goals of the United Nations is 'until 2030, to progressively reach and sustain an income growth of the 40 per cent poorest of the population at a larger rate than the national average'. Although 2015 saw a strong fall in the income of the poorest, when comparing 2007 and 2015, we observe that Brazil moved towards this goal. According to the PNAD, the income of the 40 per cent poorest among adults aged 20 years or more (including those without an income) increased at an average rate of 5.1 per cent per year in real terms, more than the PNAD average (1.7 per cent) and even more than the average in the PNAD-PIT database (2.9 per cent). The 40 per cent poorest increased their share in total income from 5.9 per cent to 7.7 per cent in PNAD, and from 4.7 per cent to 5.5 per cent in PNAD-PIT.

Table 5 details the values of minimum, average, and maximum income received in 2007 and 2015 by each tenth of distribution and by subgroups of the last tenth, the richest 5 per cent, 1 per cent, 0.1 per cent, and 0.01 per cent. The maximum value in the PNAD-PIT database in each year corresponds to the value applied by interpolation to the highest personal income.

Table 5: Income distribution in PNAD, PIT and PNAD-PIT databases in 2007 and 2015

2007 (at 2015 prices)	Mean		Minimum		Maximum		**Share of total income**	
	PNAD	PIT	PNAD	PIT	PNAD	PIT	PNAD	PNAD-PIT911
Poorest 10%	0		0		0		0.0%	0.0%
2nd tenth	0		0		0		0.0%	0.0%
3rd tenth	210		0		404		1.6%	1.3%
4th tenth	578		404		631		4.3%	3.4%
5th tenth	654		631		700		4.9%	3.9%
6th tenth	804		700		904		6.0%	4.8%
7th tenth	1,068		904		1,256		8.0%	6.4%
8th tenth	1,420		1,256		1,660		10.7%	8.5%
9th tenth	2,183	1,440	1,660		2,904	2,547	16.4%	13.0%
Richest 10%	6,412	9,851	2,904	2,547	216,063	2,792,749	48.1%	58.8%
Richest 5%	9,215	16,026	4,665	4,808	216,063	2,792,749	34.6%	47.8%
Richest 1%	18,866	45,353	11,303	17,334	216,063	2,792,749	14.2%	27.1%
Richest 0.1%	43,981	185,780	28,257	69,879	216,063	2,792,749	3.3%	11.1%
Richest 0.01%	101,498	771,860	67,142	323,949	216,063	2,792,749	0.8%	4.6%
Mean	1,333	5,646						1,675

2015	Mean		Minimum		Maximum		**Share of total income**	
	PNAD	PIT	PNAD	PIT	PNAD	PIT	PNAD	PNAD-PIT911
Poorest 10%	0		0		0		0.0%	0.0%
2nd tenth	21		0		139		0.1%	0.1%
3rd tenth	374		139		685		2.5%	1.8%
4th tenth	778		685		792		5.1%	3.7%
5th tenth	829		792		886		5.5%	3.9%
6th tenth	1,025		886		1,174		6.7%	4.8%
7th tenth	1,329		1,174		1,508		8.7%	6.3%
8th tenth	1,704		1,508		1,956		11.2%	8.0%
9th tenth	2,445	1,882	1,956		3,152	3,233	16.1%	11.5%
Richest 10%	6,712	12,742	3,152	3,233	195,625	5,958,003	44.1%	59.8%
Richest 5%	9,558	20,632	4,891	6,342	195,625	5,958,003	31.4%	49.1%
Richest 1%	19,471	58,668	11,738	21,798	195,625	5,958,003	12.8%	27.9%
Richest 0.1%	44,233	255,306	29,344	85,538	195,625	5,958,003	2.9%	12.2%
Richest 0.01%	108,463	1,170,651	65,208	416,889	195,625	5,958,003	0.7%	5.6%
Mean	1,521	7,312						2,100

Real var. 2007–15 (% a.a.)	Mean		Minimum		Maximum		**Real var. 2007–15 (p.p.)**	
	PNAD	PIT	PNAD	PIT	PNAD	PIT	PNAD	PNAD-PIT911
Poorest 10%							0.0%	0.0%
2nd tenth							0.1%	0.1%
3rd tenth	7.5%				6.8%		0.9%	0.5%
4th tenth	3.8%		6.8%		2.9%		0.8%	0.2%
5th tenth	3.0%		2.9%		3.0%		0.5%	0.0%
6th tenth	3.1%		3.0%		3.3%		0.7%	0.1%
7th tenth	2.8%		3.3%		2.3%		0.7%	-0.1%
8th tenth	2.3%		2.3%		2.1%		0.5%	-0.4%
9th tenth	1.4%	3.4%	2.1%		1.0%	3.0%	-0.3%	-1.5%
Richest 10%	0.6%	3.3%	1.0%	3.0%	-1.2%	9.9%	-4.0%	1.0%
Richest 5%	0.5%	3.2%	0.6%	3.5%	-1.2%	9.9%	-3.2%	1.3%
Richest 1%	0.4%	3.3%	0.5%	2.9%	-1.2%	9.9%	-1.4%	0.9%
Richest 0.1%	0.1%	4.1%	0.5%	2.6%	-1.2%	9.9%	-0.4%	1.1%
Richest 0.01%	0.8%	5.3%	-0.4%	3.2%	-1.2%	9.9%	0.0%	1.0%
Mean	1.7%	3.3%						2.9%

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

In PNAD, the 10 per cent richest reduced their share from 48.1 per cent to 44.1 per cent of total income, but in PNAD-PIT, they increased their share from 58.8 per cent to 59.8 per cent. The 0.01 per cent richest had the same 1 percentage point gain in their share in PNAD-PIT, which increased from 4.6 per cent to 5.6 per cent of total income, overcoming the income of the 40 per cent poorest, a group 4,000 times larger. In their turn, the seventh, eighth, and ninth tenths, which together accounted for 27.8 per cent of the total income in PNAD-PIT in 2007, reduced their share by 1.9 percentage point to 25.9 per cent.

3.5 Only positive incomes

It is even possible to revisit the analysis by excluding null incomes and considering the income distribution only among adults with income. This would cancel the distributive effect of the decrease in the frequency of null incomes from 20.2 per cent to 17.7 per cent, the average growth rates would diminish, and the Gini index would increase by 0.001 in the eight-year period instead of decreasing by 0.008. The Theil-L index, which could not be defined in view of null incomes, can thus be calculated and shows an increase. The same occurs in the J-divergence that results from adding the T and L indexes of Theil, which allows a direct decomposition of inequality among income groups and can be easily obtained even for sample databases. Tables 6–10 show the main results found for positive incomes only.

Table 6: Mean income (constant R$ at 2015 prices)—only positive income values

	2007	2015	Total var.	Annual var.
PNAD	1,670	1,849	10.7%	1.3%
Fit 0.911	2,099	2,552	21.6%	2.5%

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

Table 7: Inequality (Gini)—only positive income values

	2007	2015	Total var.	Annual var.
PNAD	0.530	0.492	-0.038	-0.005
Fit 0.911	0.622	0.623	0.001	0.000

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

Table 8: Inequality (Theil-T)—only positive income values

	2007	2015	Total var.	Annual var.
PNAD	1.417	1.173	-17.2%	-2.3%
Fit 0.911	15.646	22.144	41.5%	4.4%

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

Table 9: Inequality (Theil-L)—only positive income values

	2007	2015	Total var.	Annual var.
PNAD	0.586	0.506	-13.6%	-1.8%
Fit 0.911	1.109	1.154	4.1%	0.5%

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

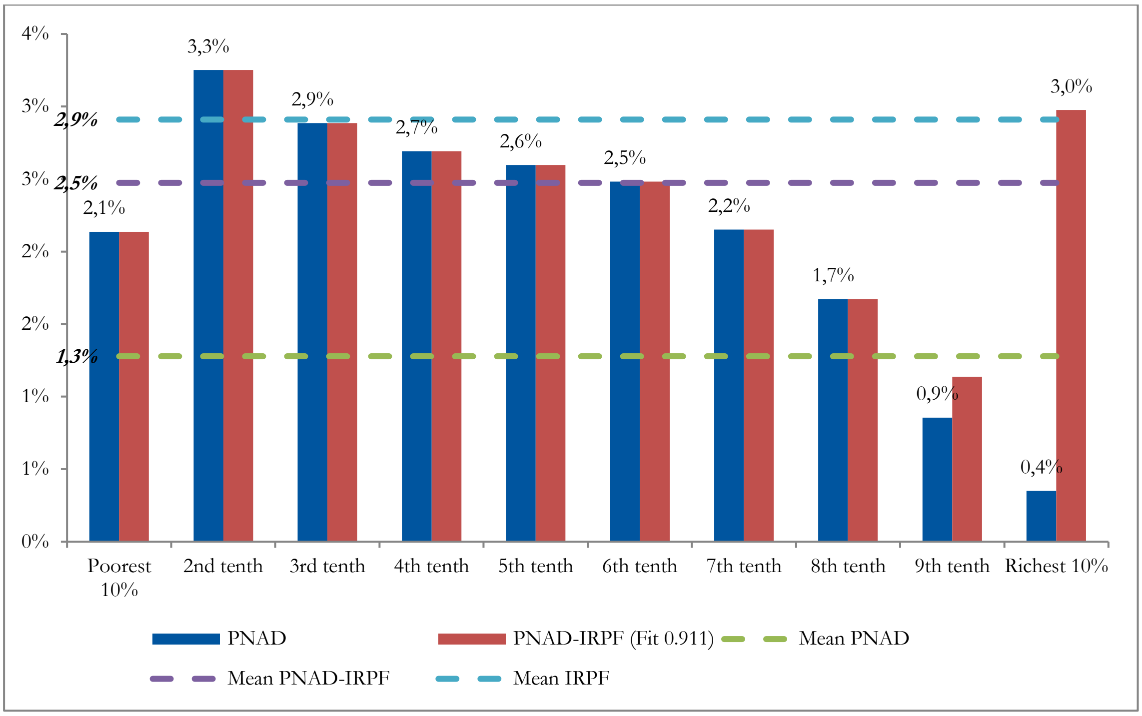
Table 10: Inequality (J-divergence)—only positive income values

	2007	2015	Total var.	Annual var.
PNAD	2.002	1.679	-16.2%	-2.2%
Fit 0.911	16.755	23.299	39.1%	4.2%

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

Figure 4 shows the increase in income for each tenth of the distribution when the null incomes are excluded. In PNAD, the poorest eight tenths experienced a larger growth than the average. In the PNAD-PIT base, the poorest tenth belong to the group with lower than average growth, together with the seventh, eighth, and ninth tenths. The greatest growth rates are in the second tenth (3.3 per cent per year) and the richest tenth (3.0 per cent).

Figure 4: Real growth per year in income by tenth of population with an income (2007–15)



Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

Table 11, besides detailing again the distribution excluding null incomes, adds a new column for the share of each group in inequality as estimated by J-divergence. The share in the J-divergence of the richest 0.1 per cent increases from 27.0 per cent to 28.8 per cent between 2007 and 2015 and is completely below that of the 40 per cent poorest, which decreases from 27.8 per cent to 26.6 per cent in the same period.

The income of the 40 per cent poorest among adults with an income increases by 2.8 per cent per year in the period, more than PNAD's averages (1.3 per cent) and the PNAD-PIT database (2.5 per cent), increasing their share in the total income in both databases. Nevertheless, their growth is smaller than that experienced by the 10 per cent richest (3.0 per cent) and their higher income subgroups.

Table 11: Income distribution in PNAD, PIT and PNAD-PIT databases in 2007 and 2015—without null incomes

2007 (at 2015 prices)	Mean		Minimum		Maximum		**Share of total income**		**Share of total inequality (% J-div)**
	PNAD	PIT	PNAD	PIT	PNAD	PIT	PNAD	PNAD-PIT911	
Poorest 10%	181		2		323		1.1%	0.9%	13.1%
2nd tenth	505		323		614		3.0%	2.4%	6.1%
3rd tenth	630		614		646		3.8%	3.0%	4.6%
4th tenth	684		646		736		4.1%	3.2%	4.1%
5th tenth	827		736		913		5.0%	3.9%	3.1%
6th tenth	1,036		913		1,165		6.2%	4.9%	2.0%
7th tenth	1,301		1,166		1,435		7.8%	6.2%	1.0%
8th tenth	1,703	255	1,435		2,017	1,294	10.2%	8.1%	0.3%
9th tenth	2,603	2,337	2,018	1,294	3,319	3,490	15.6%	12.4%	0.4%
Richest 10%	7,237	11,557	3,319	3,490	216,063	2,792,749	43.3%	54.9%	65.4%
Richest 5%	10,293	18,706	5,382	6,482	216,063	2,792,749	30.8%	44.6%	63.0%
Richest 1%	20,633	52,127	12,558	19,852	216,063	2,792,749	12.4%	24.8%	48.6%
Richest 0.1%	47,537	213,000	32,293	80,762	216,063	2,792,749	2.8%	10.1%	27.0%
Richest 0.01%	106,930	843,663	73,821	364,103	216,063	2,792,749	0.6%	4.0%	13.3%
Mean	1,670	6,947						2,099	

2015	Mean		Minimum		Maximum		**Share of total income**		**Share of total inequality (% J-div)**
	PNAD	PIT	PNAD	PIT	PNAD	PIT	PNAD	PNAD-PIT911	
Poorest 10%	215		1		391		1.2%	0.8%	12.9%
2nd tenth	652		391		778		3.5%	2.5%	5.5%
3rd tenth	791		778		792		4.3%	3.1%	4.3%
4th tenth	846		792		915		4.6%	3.3%	3.9%
5th tenth	1,016		915		1,115		5.5%	3.9%	3.0%
6th tenth	1,260		1,115		1,413		6.8%	4.9%	1.9%
7th tenth	1,542		1,413		1,660		8.3%	6.0%	1.1%
8th tenth	1,945	301	1,660		2,181	1,654	10.5%	7.6%	0.4%
9th tenth	2,786	2,867	2,181	1,654	3,519	4,463	15.1%	11.1%	0.3%
Richest 10%	7,442	14,612	3,519	4,463	195,625	5,958,003	40.3%	56.8%	66.8%
Richest 5%	10,516	23,586	5,434	8,138	195,625	5,958,003	28.4%	46.2%	64.2%
Richest 1%	21,010	66,376	13,042	24,680	195,625	5,958,003	11.4%	26.0%	49.6%
Richest 0.1%	47,413	290,752	29,344	97,722	195,625	5,958,003	2.6%	11.4%	28.7%
Richest 0.01%	115,222	1,293,925	78,250	461,700	195,625	5,958,003	0.6%	5.1%	14.7%
Mean	1,849	8,739						2,552	

Real var. 2007–15 (% a.a.)	Mean		Minimum		Maximum		**Real var. 2007–15 (pp)**		**Share of total inequality (var. pp J-div)**
	PNAD	PIT	PNAD	PIT	PNAD	PIT	PNAD	PNAD-PIT911	
Poorest 10%	2.1%		-6.1%		2.4%		0.1%	0.0%	-0.2%
2nd tenth	3.3%		2.4%		3.0%		0.5%	0.1%	-0.5%
3rd tenth	2.9%		3.0%		2.6%		0.5%	0.1%	-0.3%
4th tenth	2.7%		2.6%		2.8%		0.5%	0.0%	-0.2%
5th tenth	2.6%		2.8%		2.5%		0.5%	0.0%	-0.1%
6th tenth	2.5%		2.5%		2.4%		0.6%	0.0%	-0.1%
7th tenth	2.2%		2.4%		1.8%		0.6%	-0.2%	0.1%
8th tenth	1.7%	2.1%	1.8%		1.0%	3.1%	0.3%	-0.5%	0.1%
9th tenth	0.9%	2.6%	1.0%	3.1%	0.7%	3.1%	-0.5%	-1.3%	-0.2%
Richest 10%	0.4%	3.0%	0.7%	3.1%	-1.2%	9.9%	-3.1%	1.8%	1.4%
Richest 5%	0.3%	2.9%	0.1%	2.9%	-1.2%	9.9%	-2.4%	1.6%	1.1%
Richest 1%	0.2%	3.1%	0.5%	2.8%	-1.2%	9.9%	-1.0%	1.2%	1.0%
Richest 0.1%	0.0%	4.0%	-1.2%	2.4%	-1.2%	9.9%	-0.3%	1.2%	1.7%
Richest 0.01%	0.9%	5.5%	0.7%	3.0%	-1.2%	9.9%	0.0%	1.1%	1.5%
Mean	1.3%	2.9%						2.5%	

Source: Authors' construction from PNAD/IBGE; and combined PNAD-PIT databases.

Before concluding that inequality varied little—trusting in the method that used data from income tax—it is important to mention that the number of income tax files increased by 9.1 per cent between 2007 and 2015, while the 20+-year-old population increased by 9.5 per cent. The proportion of adults filing tax returns decreased, and the number of files with income of up to 2 MW decreased, whereas files with a higher income increased. Likewise, the number of tax files from people over 40 years old dropped, while younger filers increased. Finally, the number of dependants increased, and the average income filed increased considerably, pulled by an 'economic miracle' type of growth between 2007 and 2011.

Is it possible, then, that the PNAD has increasingly missed the income of the richest and hence missed the 'miracle' they experienced? Or is it that the income tax system became more accurate in measuring the income of a larger group, pointing to a fantastic growth precisely because it identified previously unseen incomes? It is not easy to answer these questions without accessing IRS microdata, or at least tables that show longitudinal movements of taxpayers, dependants, and people who are included in the income tax database in any condition. For now, the best recommendation is caution in the interpretation of the IRS tabulations and the attempts to integrate them with household surveys. The use of tax data disclosed in recent years enables new hypotheses but does not by itself warrant definite conclusions on how income inequality and income distribution as a whole have varied in Brazil.

4 The 'economic miracle' in the personal income tax tables

To understand the combined PIT-PNAD distribution levels and changes, related features and sources, one should also analyse PIT data in isolation. PIT figures as released by the IRS are available in standard format for the 2007–15 period. PIT mean income growth for the 2007–15 period presents astonishing differences with respect to GDP growth: 4.97 per cent per year against 1.23 per cent. That is a 304 per cent faster growth rate, or a gap of 3.74 percentage points per year (ppy). If we take the PIT income data at face value, this difference yields non-negligible impacts on social welfare growth in the period. What is perhaps more relevant is that this gap is worth investigating because it may allow us to track sources of measurement error that affect not only mean income growth, but also inequality changes, since they all come from the same information set.

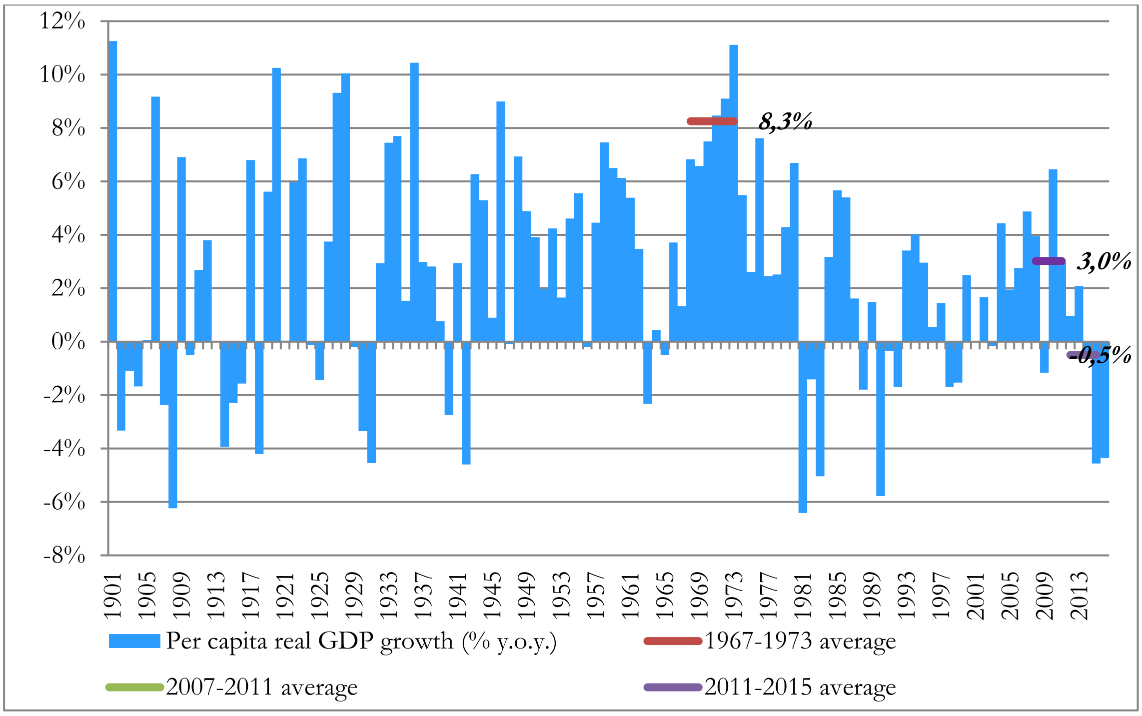
4.1 Mean income growth

For the same reason, it is worth breaking the analysis into different periods. The real growth rate of the mean income per filer between 2007 and 2011 is impressive, i.e. 10.1 per cent per year when applying the Broad National Consumer Price Index (IPCA) as deflator. During the same period, the IBGE reveals that per capita GDP grew 3 per cent per year (Figure 5).

Between 1967 and 1973, Brazil experienced the so-called economic miracle, with high annual growth rates. In that period, per capita GDP grew by 8.3 per cent per year on average, as Figure 5 shows. Since the beginning of last century, per capita GDP has increased by more than 10 per cent only in five years: 1901, 1920, 1928, 1936, and 1973, the last year of the 'miracle'.

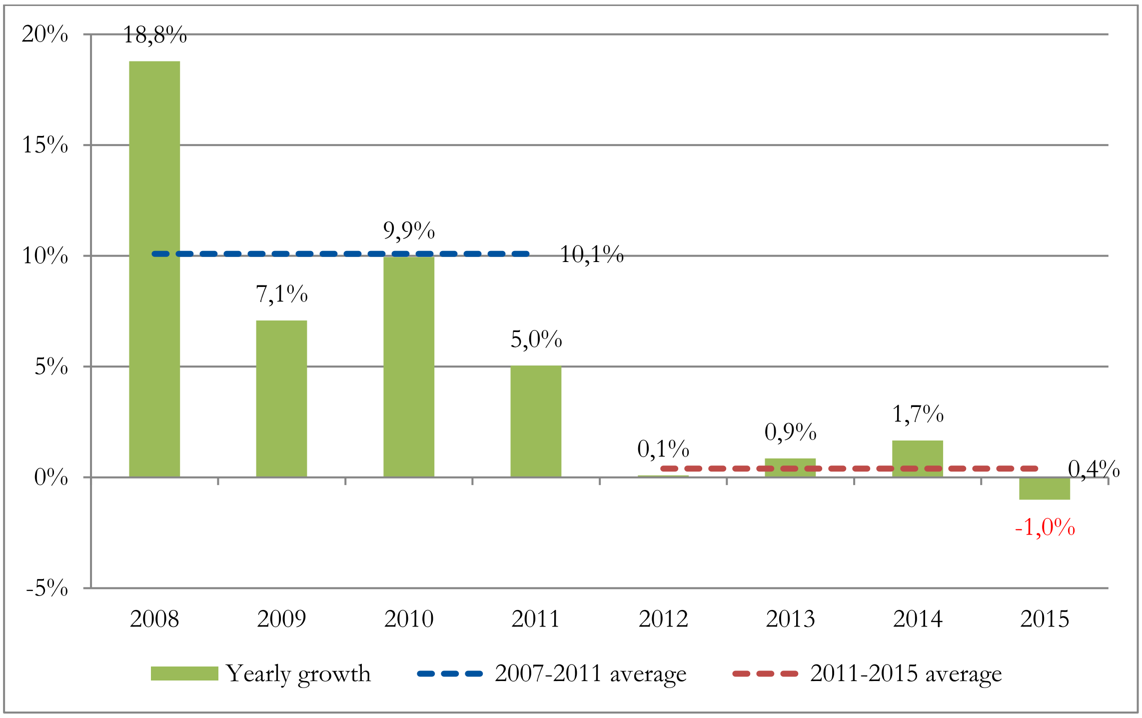
In the PIT, the average income per file increased by 18.8 per cent in 2008, 7.1 per cent in 2009, 9.9 per cent in 2010, and 5.0 per cent in 2011 (Figure 6). In the next four years, growth in the PIT was milder, about 0.4 per cent per year on average, closing the series with a decrease of 1.0 per cent in 2015, when per capita GDP dropped by 4.6 per cent.

Figure 5: Real growth rate of per capita GDP in Brazil—1901–2016 (% a.a.)



Source: Authors' construction from IBGE and IPEA data.

Figure 6: Real growth rates of mean income per PIT file (% y.o.y.)



Source: Authors' construction from IRS. Deflation by IPCA/IBGE.

All the calculations done so far take into account real mean PIT per declarant but there was also a rise in the number of declarants from 25.22 million in 2007 to 27.52 million in 2015, a growth rate of 1.1 ppy in the period, widening the gap to be explained to 4.88 ppy.

4.2 Propensity to declare

One line of reasoning has to do with the increasing formalization of the Brazilian economy during the 2007 to 2015 period, which may lead to an overestimation of the mean income growth of those that declare PIT. The idea here is that, as formalization progressed, the Brazilian IRS became more able to observe incomes, which creates the impression that incomes were growing faster than they were in reality. Therefore, the growth of income in the PIT encompasses both actual real income growth and the increasing ability to see that. This second component can be captured by the rise in the share of the occupied population that contributes to social security in the PNAD. It rose from 50.3 per cent in 2007 to 61.52 per cent in 2015, an annual growth rate of 2.56 per cent. That could explain more than half of the gap of 4.88 ppy, still leaving 2.26 ppy to be explained.[4]

4.3 Deflators

Another part of the explanation of the remaining gap is related to the differences in the price index used. The GDP implicit deflator from the National Accounts grew 1.71 ppy faster than the IPCA.[5] If we use the same price deflator in both nominal income series, this also corresponds to the nominal growth rates differential. One advantage of this reasoning is to leave both nominal National Accounts and PIT records untouched while filling part of the puzzle. Neri (2009, 2014) has shown that using the same deflator also allows us to reconcile almost all differences between GDP and standard PNAD income growth differences in the 2003–13 period. These annual differences turn out quite similar to the 2007–15 period. Applying the IPCA to nominal GDP instead the usual implicit GDP deflator, the observed real growth gap would fall from 2.26 ppy to 0.54 ppy.

Looking into the data by type of income allows us to progress towards a growth gap puzzle solution, which will be done in detail in the next section. Shortening a long story, we captured an overestimation of financial gains growth that amounted to an additional impact of 0.35 ppy. The remaining gap of 0.19 ppy to be explained amounts to a small share of 3.86 per cent of the original growth gap between GDP and PIT.

5 Inspecting income types: the growth of exempt sources

5.1 Overview

The use of income tax data to adjust for estimates of the income distribution in Brazil assumes that the people who filed their tax returns earn at least what they declared to the IRS (Morgan 2017: 3). It is true that it is not incentive-compatible in general to declare higher taxable income than is actually earned, because it leads to higher taxes. But the argument does not apply to non-taxable income sources. As a result, any increase in the share of non-taxable income potentially challenges the initial idea that led to the replacement of PNAD data by PIT records on the top incomes range.

---

[4] One point that needs further elaboration is why the number of declarations (1.1 ppy) increased less than formalization rates (2.56 ppy). Section 6 addresses this question.

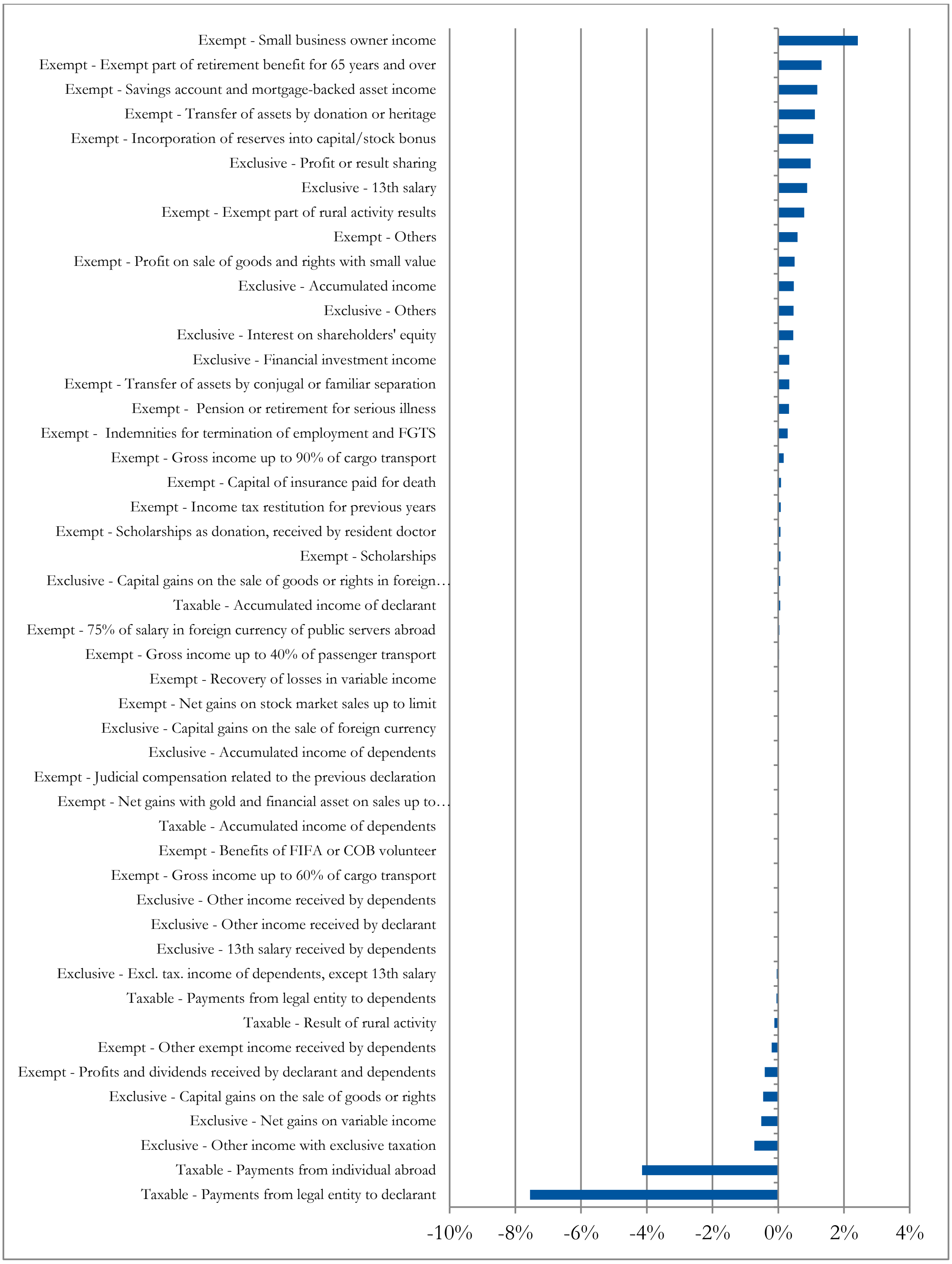
[5] In the 2007 to 2015 period the GDP implicit deflator at market prices grew 7.96 per cent against 6.14 per cent of the CPI (IPCA). One may argue that the latter is better suited to measure social welfare changes.

From 2007 to 2015, exempt and non-taxable income grew at an average annual rate of 11.45 per cent, against 3.88 per cent from taxable income. In the 2007–11 period this difference was even higher: 18.74 per cent against 5.45 per cent. Finally, even more distant are changes observed between 2007 and 2008: 62.87 per cent against 6.82 per cent, respectively. A detailed look at time periods and more specific income sources can clarify institutional causes and distributive consequences behind the profile of incomes declared.

As a result of these differential growth rates, in the whole 2007–15 period, exempt and non-taxable income share increased from 21.4 per cent to 31.3 per cent (+9.9 percentage points) of the total income filed in the IRS, while taxable revenues reduced their share from 70.7 per cent to 58.9 per cent (-11.8 pp) and income subject to exclusive tax increased its share from 7.9 per cent to 9.8 per cent (+1.9 pp). To have a more complete picture, one has to infer the reasons behind this dramatic increase in the share of non-taxable income.

Figure 7 zooms in on the changes in the participation of the 48 sources that make up these groups of income. The source that lost the most of its share—in other words, the one which grew the least—was the amount received by taxpayers from employers/companies, the largest source of taxable revenue of all. The share of this source dropped 7.6 pp from 60.8 per cent to 53.2 per cent. The income sources that grew the most were all exempt, namely: small and microenterprises (+2.4 pp), the exempt share of retirement income of people aged 65+ years (+1.3 pp), and savings account income and mortgage debts (+1.2 pp).

Figure 7: Change in the share of each income source in the total tax return filed from 2007 to 2015 (pp)



Source: Authors' construction from PIT/IRS.

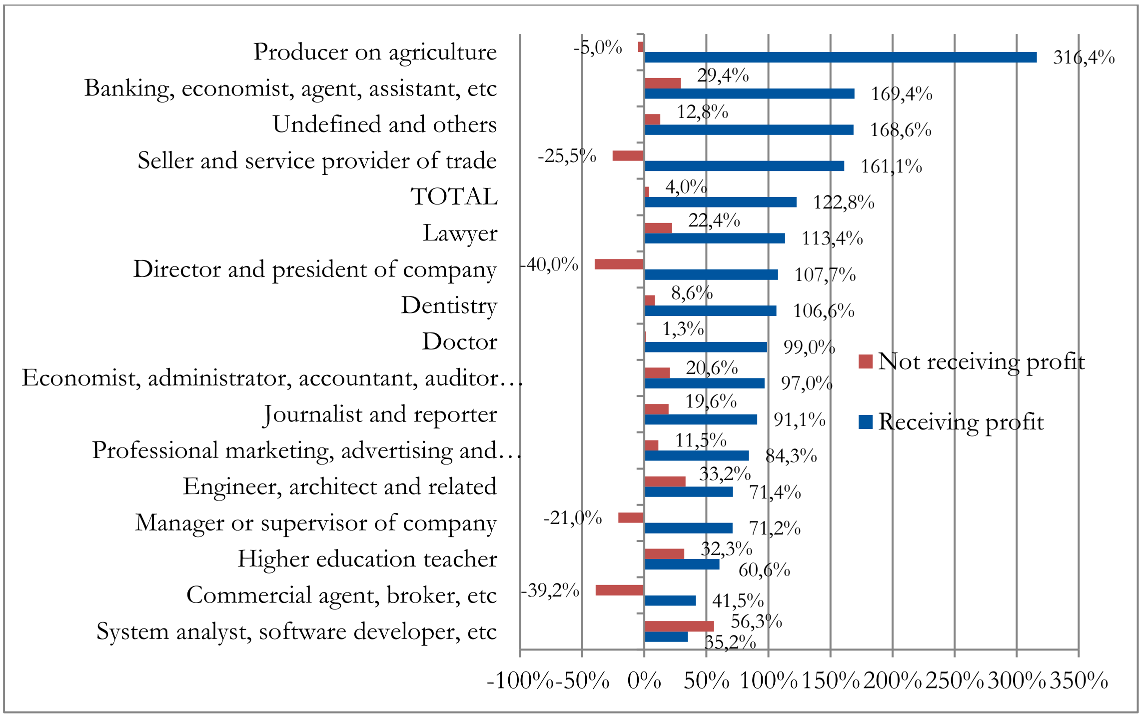
5.2 Small business formalization and transfiguration

The growth of exempt small business owner income may relate to an increase in their profits, but the main cause may be the greater formalization of businesses hitherto not detected by the PIT data alone. This is a specific item applied to small business of the formalization argument mentioned above. Alternatively, it may be due to a growing process of transfiguration (Afonso

2017) of individuals into legal entities, that is, the hiring of workers as microenterprises to bypass the costs imposed by the Brazilian labour law. In this last case, the declarants would not be capitalists perceiving greater income from their relationship with workers, but workers who present themselves as companies to the IRS. This result is consistent with the greatest fall of the 48 income sources pointed to above, which is a taxable income item: payments from legal entity to declarant, which fell 7.6 pp in the 2007–15 period.

Figure 8 displays the variation in the number of files with profit income or not by the main occupation of the filer. Farmers constituted the greatest leap in the number of files that registered income from profit (+316.4 per cent), just as they reduced (-5.0 per cent) their files with no profit income. The smallest growth rate in the number of filers with profit income related to system analysts and others (+35.2 per cent), who also had the largest growth in the number of files without profit income.

Figure 8: Variation 2007–15 of files with profits income, or not, by main occupation (%)



Source: Authors' construction from PIT/IRS.

The number of files from receivers of profits and dividends, partners, or owners of microenterprises increased 122.8 per cent between 2007 and 2015 (10.5 per cent per year), increasing their share in the total number of files by 4.5 pp, from 4.3 per cent to 8.8 per cent. In the same period, the share of employers and self-employed according to the PNAD grew by only 0.44 per cent per year.[6] At the same time, the greater formalization of businesses—a 12.4 per cent per year rise according to PNAD 2007–15 as a result of the 2009 MEI (individual microentrepreneur) Law—also led to the transfiguration of employees into self-employed (Corseuil et al. 2014).

[6] The share of employers according to the PNAD remained stable at 3.7 per cent, while the share of self-employed grew by 1.7 pp from 21.2 per cent to 22.9 per cent.

5.3 Exempt retirement benefits

The second source of expansion in the exempt income group is due to the increase in the volume of pension benefits and the growing share of these benefits that may be exempt from income tax. This is the result of a combined increase in the concentration of benefits with lower values of up to 2 MW, and the growing number of older people receiving these benefits (MF/DATAPREV 2016).

This phenomenon interacts with the one scrutinized in Section 6.3, which reports a reduction in the number of elderly declarants and their possible re-allocation as dependants of their offspring (or even grandchildren). The institutional change that apparently triggered this movement is also presented there. For now it suffices to consider this supposed conversion of older taxpayers into dependants, which could at least partially explain the drop of 3.1 million in the number of tax return files among those aged 41[7] years or more between 2007 and 2015. The taxpayer population aged 41+ decreased by 15.9 per cent in the period, from 19.7 to 16.6 million, while according to the PNAD it grew by 30.3 per cent. At the same time, it doubled the average number of dependants declared by each filer up to 40 years of age, from 1 to 2 dependants for every 3 files within this age group of filers. A growing share of people aged 41 and above registered as economic dependants would help explain the intriguing decrease in the number of 41+-year-olds filing their tax returns, in opposition to the prevailing demographic shift in the overall population.

It is also true that there was strong increase in the share of people up to 41 years old who file income tax returns, as well as a relative gain in the income of younger workers, concomitant with a reduction in the return on experience. According to Ferreira et al. (2017), this was the main determinant of the decrease in labour earnings inequality between 1995 and 2012. These authors suggest two possible explanations for this phenomenon. The first is an age-biased technical change resulting from a greater demand for younger workers, who are likely to be more productive with new technologies. The second is a shift in Brazilian production towards those sectors benefitting from changes in the terms of trade that occurred in the period, favouring younger workers flexible enough to enter more dynamic sectors. This would reduce the premium on the experience of older workers in those sectors that lost their relative importance.

5.4 Financial income overestimation

Finally, the third source mentioned, savings account income and mortgage debts, is highly susceptible to overestimation of financial investment income in the PIT files (Hoffmann 2017), a likelihood that was aggravated between 2007 and 2015. The period saw increases in specific non-taxable incomes. This is consistent with the overestimation of the income processed in the IRS tables because it uses information on nominal instead of real interest incomes. Indexing a financial asset according to inflation rates only prevents it from losing its purchase power. It should not be considered as income. However, what Brazilian banks declare as 'income' from financial investments, and communicate to the IRS as such, is the flow of nominal interests, which is equivalent to the real interests combined with indexation. Including this 'fictitious income'—based on 'monetary illusion', as Hoffmann has put it (2017: 386)—may explain the positive correlation that he points out between the Brazilian annual inflation rate and the inequality indexes reported by Medeiros and Souza (2015) and Morgan (2015) based on PIT files data. These correlations overcome 0.8 and are statistically significant at 1 per cent, even though there have been only eight observations in the series analysed by the author.

[7] According to PNAD 2007, 41 corresponded to the peak of occupation in a static age profile.

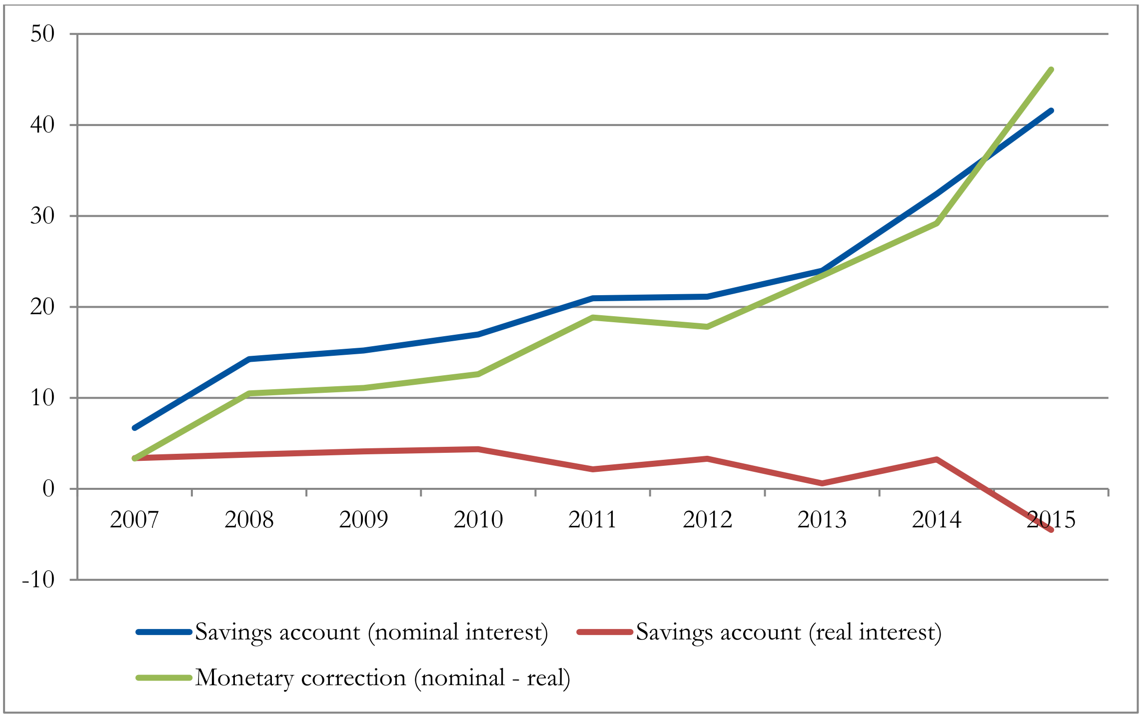
To estimate the real income from interest equivalent to the nominal income presented in the IRS tables, it is not enough to deflate the filed values. Besides the inflation factor, it is necessary to know what is the amount corresponding to the flow of interests or the profitability rate of each investment. IRS tables show total values, on 31 December each year, of the total assets as filed, but as this information is not used to calculate the tax, it may be omitted more frequently or erratically than the information on the flow of income from interests received each year.

We have opted to estimate the effect of part of this 'fictitious income' based on the average annual profitability rates of two types of financial investments informed by the Brazilian Central Bank (BCB): savings account and bank deposit certificate (CDB).

The IRS table of exempt or non-taxable income includes the total amount filed annually in terms of savings account and mortgage debt income (R$41.6 billion in 2015). This total contains all income-tax-exempt investments and therefore mixes the income from Brazil's most popular financial asset, the savings account, with the income from mortgage debt, which is 600 times less prevalent nationally.[8] For that reason, in the following exercise, mortgage debt was ignored and only the average profitability of saving accounts was used to estimate the asset that corresponds to the flow of nominal income and its respective indexation and real income.

In the case of this asset's profitability, Figure 9 shows that the nominal interest rate varied little between 2007 and 2015, but inflation measured according to the IPCA grew; thus the real interest rate dropped in the same period, reaching negative levels in 2015.

Figure 9: Average nominal and real profitability of savings accounts and inflation rates



Sources: Authors' construction from BCB (nominal interest) and IBGE (IPCA deflator).

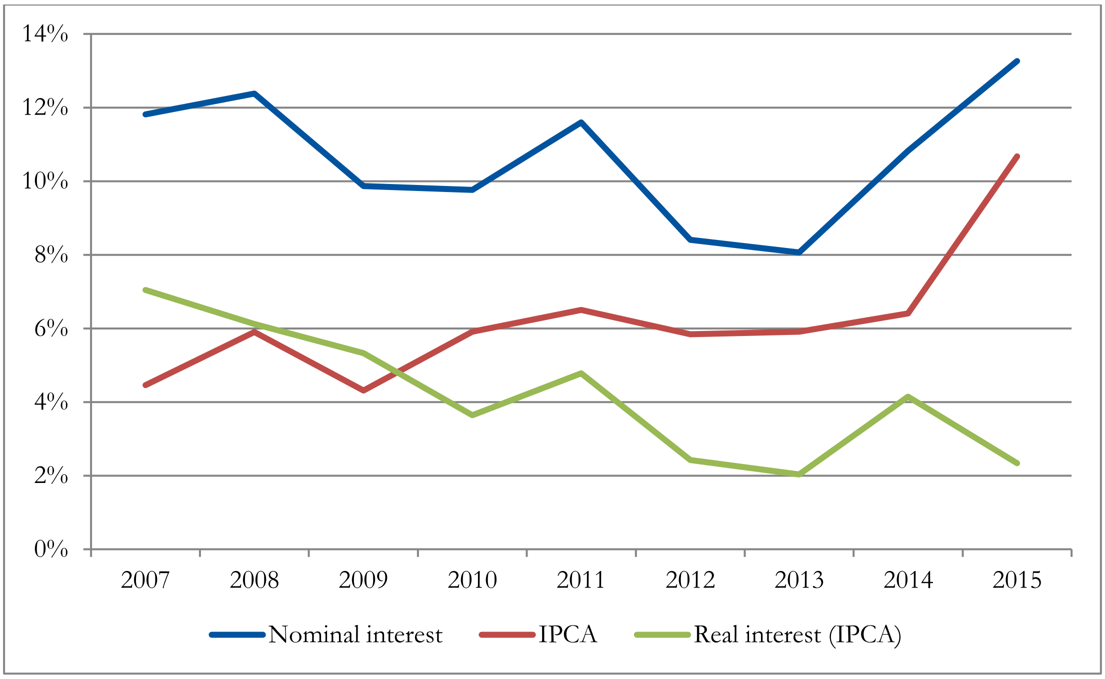
[8] In June 2017, savings account balances totalled R$667.6 billion, distributed among 149.5 million clients, while mortgage bills totalled R$1.1 billion, with 391 clients.

Income tax return files do not compute negative income, but the 'monetary illusion' is not restricted to these cases. Figure 9 shows that the income from savings account interest displayed in the income tax tables increased from 2007 to 2015 following our estimate of the indexation flow, while the real income that savers received decreased from 2011 until becoming negative in 2015, when the nominal rate was lower than inflation. The green line, referring to the monetary indexation, illustrates the estimated size of the 'fictitious income' from savings accounts at 2015 prices[9], which leaps from R$3.3 billion to R$46.1 billion in the period analysed. This real growth of 1,285 per cent in the flow of monetary indexation explains the contribution of savings account and mortgage debt income to the already mentioned expansion of exempt and non-taxable income.

Regarding the income to be exclusively taxed, which is always paid at source, the second highest figure in the IRS tables is the so-called income from financial investments (R$69.7 billion in 2015). This may include different types of assets with varied profitability, but excludes other operations, such as net gains in variable income (R$3.7 billion in 2015). To estimate the 'monetary illusion' relating to the income from financial investments, we used the average profitability of the CDB, a private bond for fixed-term deposits.

Figures 10 and 11 show the results. First, the accelerated inflation causes the CDBs' real interests to drop between 2007 and 2015. Next, the nominal flow of income from financial investments grows in the IRS tables, but this movement is caused by an increase in the flow of monetary indexation, whilst the estimated real income decreases. The flow of monetary indexation at 2015 prices, which may be interpreted as 'fictitious income', increases by 296 per cent in real terms, from R$12.4 billion in 2007 to R$49.3 billion in 2015.

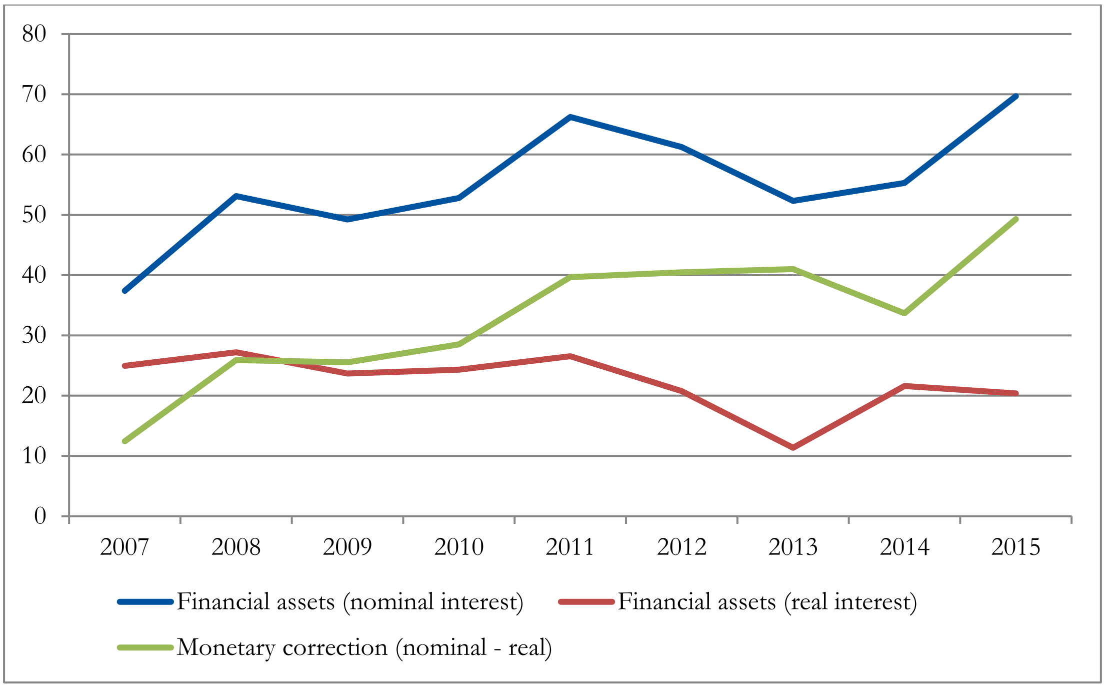
Figure 10: CDBs' average nominal and real profitability and inflation rates



Source: Authors' construction from BCB (nominal interest rate) and IBGE (IPCA deflator) data.

[9] The IPCA level in December each year was used to deflate the stocks, and the average IPCA level for the 12 months of each year was used to deflate the flows.

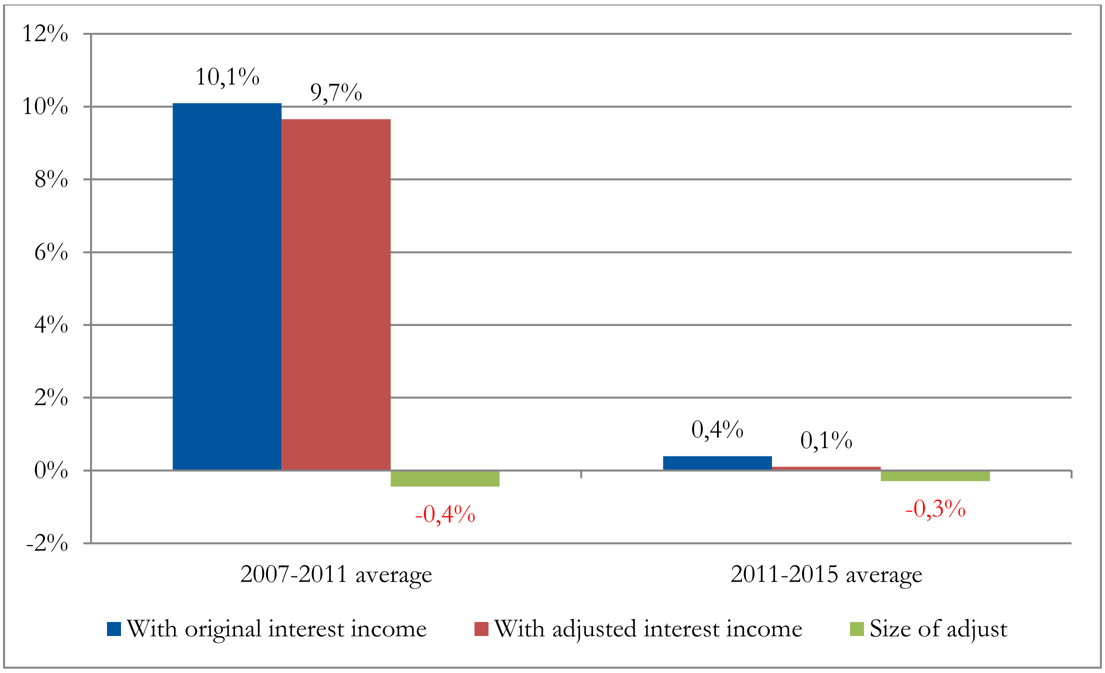
Figure 11: Nominal and real income from financial investments and monetary indexation (R$ million in 2015)



Sources: Authors' construction from IRS (nominal interest) data and estimates based on BCB and IBGE data.

If the series on total income is adjusted to exclude these estimates of monetary indexation of savings accounts and financial investments, the average annual growth rate of total income per file would be attenuated by 0.4 pp in the 2007–11 period and by 0.3 pp in the 2011–15 period. Figure 12 shows that, even after this adjustment, the 2007–11 period still has a 9.7 per cent per year real growth rate of average income per file. The 'economic miracle' in the income tax tables remains, even if the 'monetary illusion' of these two financial assets is excluded.

Figure 12: Real growth rate of average income per income tax file (% annual)



Source: Authors' estimates based on IRS data.

6 Taxpayer's profile challenges demographic patterns

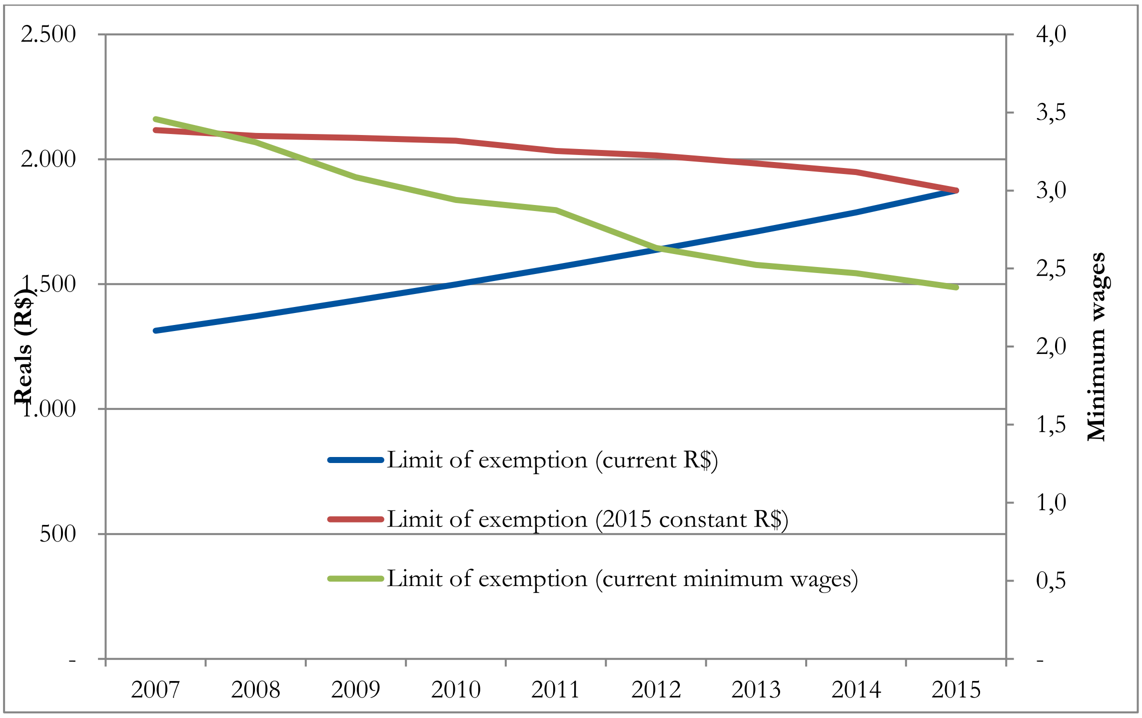
The rationale of PIT analysis must also take into account the existence of incentives that make conclusions taken based on PIT data—or on the combination of PIT records and household surveys—more complex. These incentives can affect the number of files, the age distribution of filers, and the number of their dependants, whose trends should be studied.

6.1 Number of files

The simple evolution in the number of files processed by the IRS is puzzling. Between 2007 and 2011, the country's per capita GDP grew, the population grew, more formal jobs were created, but conversely the number of return files decreased by 1.2 per cent from 25.2 million to 24.9 million. From 2011 to 2015, however, the number of files increased by 10.5 per cent, reaching 27.5 million, despite the economic slowdown. For the whole period between 2007 and 2015 the number of files grew by 1.1 per cent annually on average.

Figure 13 shows that for every year in this period, the limit of income tax exemption in current values increased in a linear fashion, at 4.5 per cent per year. The IPCA inflation rate was completely below this nominal adjustment each year, reducing the real value of the exemption cap year after year. The minimum wage increased in real terms (above inflation); hence, the exemption cap dropped further in terms of minimum wages. In 2007, monthly income of up to 3.37 MW was exempt from income tax; in 2015, only monthly income below 2.38 MW was exempt.

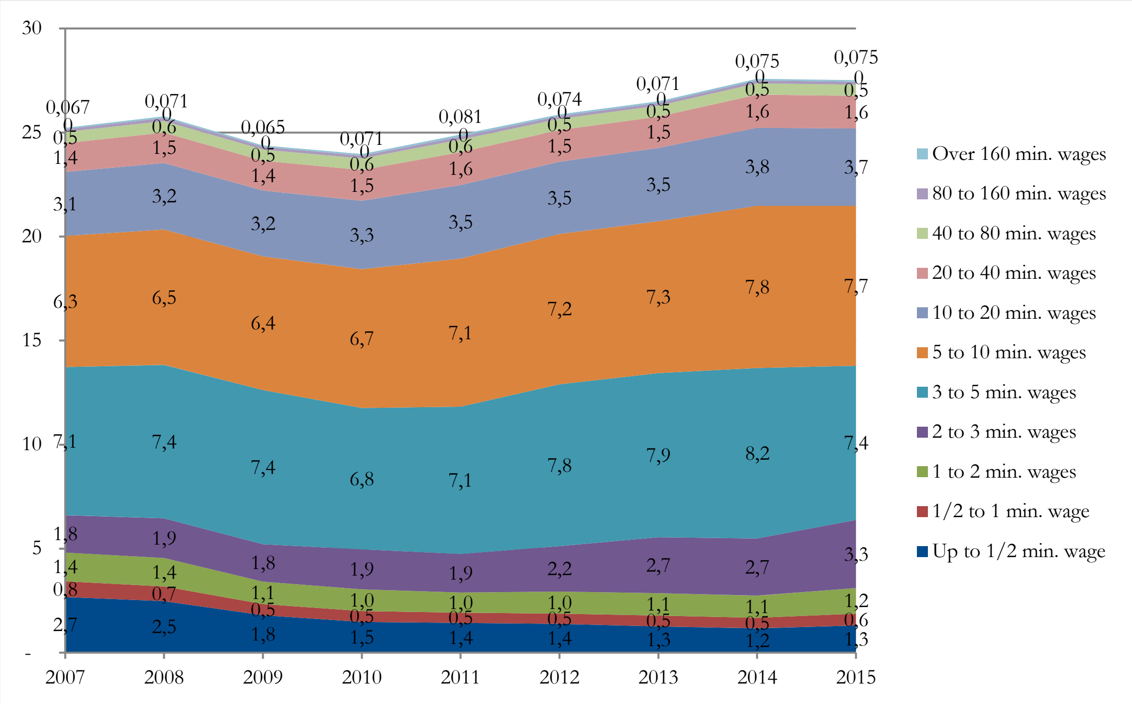
Figure 13: Income tax exemption cap—monthly income



Source : Authors' construction from IRS data. IPCA deflator.

Since the exemption cap decreased in real terms, a greater number of files would be expected, especially in those years of greater economic growth. Nevertheless, as seen, the number of files decreased in the period of greater economic growth. Moreover, as Figure 14 shows, the number of files for total income of up to 3 MW did not increase between 2007 and 2015, although this group benefitted from the reduced exemption cap in the period. In fact, this number diminished from 6.6 million to 6.4 million, while the number of files for total income above 3 MW increased from 18.6 to 21.1 million.

Figure 14: Distribution of files according to income level as presented in the IRS tables (million)



Source: Authors' construction from PIT/IRS.

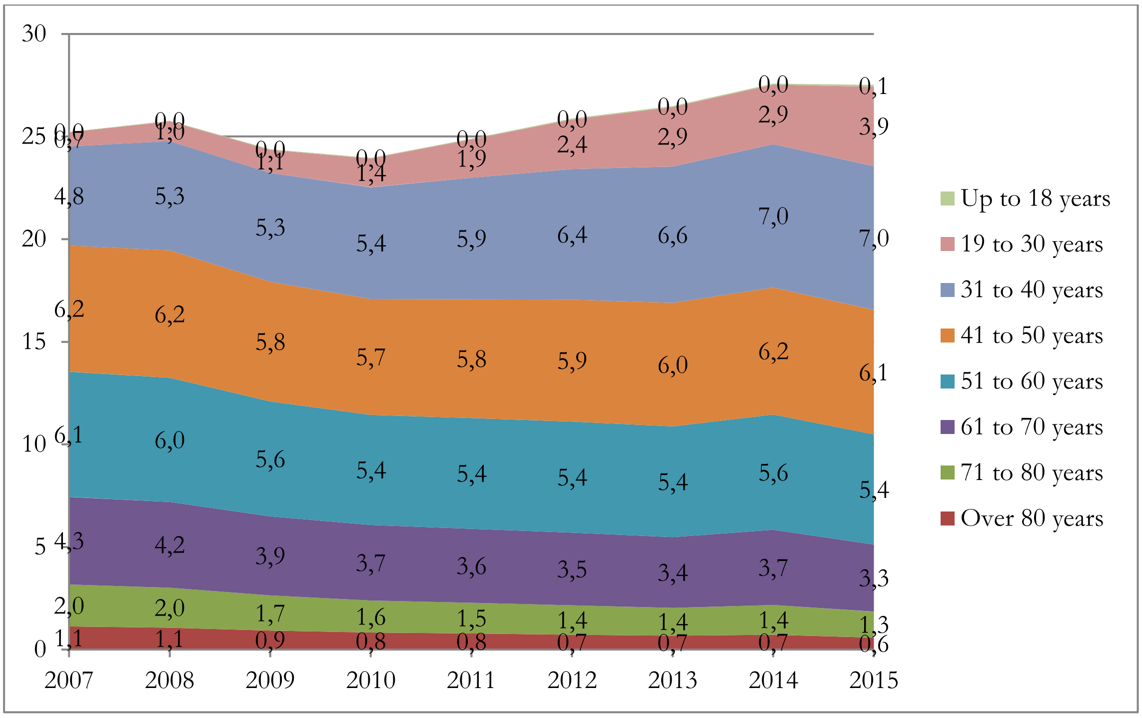
The number of files of income between 2 and 3 MW grew the most, both in relative terms (+83.2 per cent) and in absolute terms (+1.5 million files). Nonetheless, the number of files for total income below 2 MW dropped by 1.7 million, with decreases of 51.4 per cent for the group earning up to ½ MW, 24.0 per cent for those earning between ½ and 1 MW, and 11.5 per cent from 1 to 2 MW. At the same time, at the other extreme, tax return files for incomes above 160 MW increased by 13.2 per cent, from 66,600 to 75,400.

The fundamental hypothesis underlying all work that uses the income tax tables to track the evolution of income distribution is that, each year, the income tax files containing the highest figures correspond to those with the highest income among the population. Nevertheless, it is possible that income not filed varies over time and that the mix of population varies. Some of these changes may be inverse to what is expected from the economic or populational dynamics, as other well-known data sources will show next.

6.2 Age

Although the Brazilian population is aging, income taxpayers are getting younger. From 2007 to 2015, taxpayers aged 41 or more decreased not only in their share in the total tax return files, but also in their absolute number, which fell by 15.9 per cent in the period, from 19.7 million to 16.6 million. The IBGE estimates that the 41+ group increased by 25.1 per cent in the same period, from 55.3 million to 69.2 million people, while the 15–40-year-old group grew only by 0.7 per cent. Nevertheless, the total number of processed tax return files increased (+9.1 per cent) between 2007 and 2015, from 25.2 million to 27.5 million. As Figure 15 shows, the decrease in the number of 41+-year-olds filing tax returns was outweighed by the 98.0 per cent boom in the number of taxpayers up to 40 years old, from 5.5 million to 11.0 million.

Figure 15: Number of income tax files per age group (millions)



Source: Authors' construction from PIT/IRS.

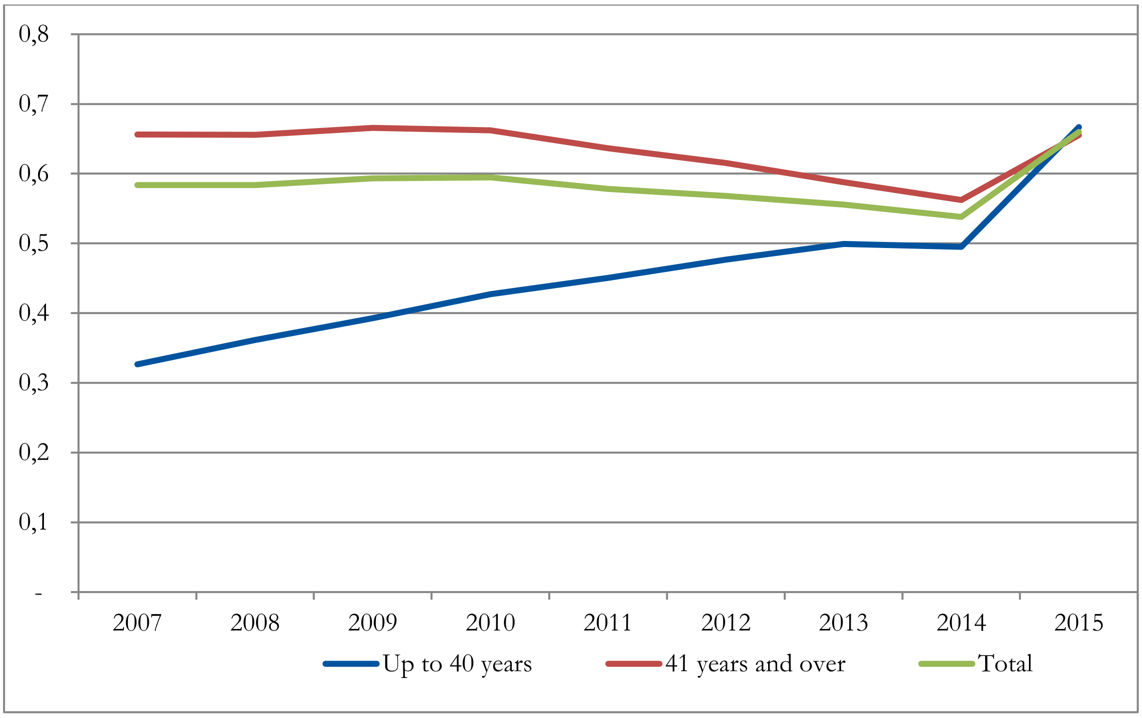
6.3 Dependants

The income pertaining to each file does not reflect the income of a single person, but the income of the person who filed the tax return plus that of his/her dependants. Changes in the number of dependants per file and their allocation to other groups may bias the analysis, and there is enough evidence to suggest that there were crucial changes during the period of interest. The IRS tables do not display the number of dependants for each group of filers, but they show the total monetary amount of deductions with dependants in each group. This can be divided by the annual legal amount of deductions for each dependant, whose historical series are available at the RFB website[10], to provide a proxy estimate of the number of dependants in each group of filers.

Using this estimate, Figure 16 shows that the average number of dependants per filer up to 40 years of age doubled from 1 to 2 dependants for every 3 files within this group, matching the average for those over 41 years old. A growing share of people aged 41+ registered as economic dependants would help explain the intriguing decrease in the number of 41+-year-olds filing their tax returns, in opposition to the demographic shift, but does this hypothesis make sense?

[10] As of January 2018: http://idg.receita.fazenda.gov.br/acesso-rapido/tributos/irpf-imposto-de-renda-pessoa-fisica#dedu--o-anual-por-dependente.

Figure 16: Estimated average number of dependants according to the age of main taxpayer



Source: Authors' construction from PIT/IRS.

It should be noted that, according to the PNAD, the average number of inhabitants in each household in Brazil decreased from 3.36 in 2007 to 3 in 2015. However, dependants of a person who files a tax return do not necessarily live in the same household.

6.4 Rationale

Someone aged 41 years or more could feature as someone else's dependant by various statuses—partner, spouse, parent, grandparent, or in-law—as long as their revenue did not reach a certain level, which in 2017 was R$1,903.98 monthly. This procedure also allows us to discount dependants' health expenses. Taxpayers, who may file their taxes together or separately, will chose the more advantageous option. In the case of the over 40s, the joint tax return file is a better option when the largest share of person's income is exempt from tax, thus preventing the joint file from incurring higher tax rates. It could also allow a deduction in the total tax due, as the dependant may have medical or other expenses that would have no effect in the case of exempt income.

Until 2008, each declarant had to complete a PIT form in order to have a valid Social Security number (CPF) in Brazil. This was a way to try to control income tax evasion in the country. In 2008, this obligation was dropped, which created an extra incentive to move from declarant to dependant status in the PIT records if the dependant is in the relevant income range.[11] Furthermore, Brazilian income tax legislation allows the individual to include as dependants both

[11] On 25 July 2008, the IRS introduced a norm (Instrução Normativa 864) that abolished the requirement to make an Annual Declaration of Tax-exempt Income (Declaração Anual de Isento). In 2007, this exemption ceiling was R$15,700 in taxable income or R$40,000 for total exempt income or other income sources. In 2007, there were 24 million declarations and 66 million exempt declarations, which corresponded to around 90 million active social security numbers (CPFs). On top of that, there were 48 million people with suspended numbers or waiting for regularization or their numbers.

parents (and the parents of their parents and so on). So, if a couple opts for a joint declaration, it can include all living parents, grandparents, great-grandparents, and so on.

Our main hypothesis here is that the movement of the elderly from declarant to dependant has been influenced by the imposition of this new norm. Financial incentives and, perhaps more importantly, the reduction of transaction costs could explain the process of including more income sources within the same PIT declaration. This line of reasoning has the potential to explain a few changes that occurred between 2007 and 2008 (as well as lagged effects of these changes that might have lasted until 2011, including the yearly growth of the 2007–11 period), namely: (i) the sharp rise in mean declared real income of 18.8 per cent (10.1 per cent); (ii) the relative fall in the number of declarations by the elderly of -1.23 per cent (-3.5 per cent) with respect to the formalized population; (iii) the increase in the number of dependants of people below 41 years of age of 9.1 per cent (36.4 per cent).

7 Conclusions

Recent studies combining data from household surveys with administrative records of personal income taxes hold the promise of measuring more accurately top incomes. This literature emphasizes the rising income inequality impact of this data-combination exercise, but it does not detail the effect on mean income or social welfare. If the level of inequality measured rises when higher top incomes replace previous lower estimates based on surveys, this same exercise also increases the mean and the social welfare level associated with it. The implications of these combined estimates in terms of the movement of these series across time is an empirical matter.

Illustrating this initial point, the pioneering work by Medeiros et al. (2015a, 2015b) applied this type of approach to Brazil between 2006 and 2012. It indeed showed that inequality as measured by the Gini index was 11 per cent higher in the final year and decreased by 2 percentage points less during this period than indicated by the traditional PNAD-based estimates. Conversely, our calculations on their published statistics reveals that Brazilian income would also be 35 per cent higher and would have grown 13 percentage points more between 2006 and 2012. If we were to combine these qualitative results with impacts on social welfare by using Amartya Sen's (1974) specification, the level of the combination of mean income with inequality as measured by the Gini index on a synthetic index of social welfare would be 9.62 per cent higher and would have grown by 11 percentage points more in the analysed period. However, more generally, the growth incidence curve comparing the two extreme years is always in the positive quadrant, showing an unequivocal increase in social welfare—in fact a Pareto improvement—during this period.

The present paper has aimed to evaluate the implications of combining household surveys with income tax return files in Brazil from 2007 to 2015. We extrapolated the impacts on income inequality, also incorporating mean income and social welfare into the picture. In this regard, we suggested several extensions to previous literature. First, we included the period after 2012, when a major Brazilian recession began, challenging the welfare improvement mentioned. Second, we tested different fittings in the distributions and evaluated their implications by using different aggregate measures of welfare and their components, including income mean and inequality indexes. Last and not least important, we attempted to address jointly the reasons for the static and the dynamic behaviour in the two first moments of the income distribution.

We considered the population aged 20 or more and their individual income from all sources. We performed a series of robustness tests on the compatibility of both survey populations and explored different income quantile fittings. The analysis described here used the same fitting point in both years (0.911 quantile). We referred again to the social welfare measurement criterion proposed by Sen (1974), which results from multiplying mean income by the Gini inequality index complement. The welfare level in 2015 was 2.36 per cent higher for the combined distribution vis-à-vis the PNAD alone. Although the Gini was 18.6 per cent higher, the mean income was also 38.5 per cent higher in the same year. The welfare gain between 2007 and 2015 in the combined database is 3.2 per cent per year. The same measure using only PNAD was 3.0 per cent per year. This superior performance of the country according to Sen's measure happens because—despite inequality measured by the Gini index having decreased 0.7 pp less in the combined database—income growth was 1.2 pp higher each year. The inequality result varies with the different measures used because the Lorenz curves for 2007 and 2015 do cross, while the corresponding ones for PNAD do not. Therefore, by this criterion, looking jointly at both shifts in the first two moments of income distribution expressed as welfare, growth is higher in the combined database than in the PNAD alone.

Among all the differences between PIT and PNAD, the faster income growth trends of the former are the main driver behind the results mentioned above. This was an object of detailed investigation here. PIT income tax tables present even more astonishing differences with respect to per capita GDP growth: 4.97 percentage points per year (ppy) against 1.23 ppy. That is a gap of 3.74 ppy. If we take into account the rise of 1.1 ppy in the number of tax declarants, the gap to be explained increases to 4.88 ppy.

To explain this gap, we initially combined two ingredients. The first is the increasing formalization of the Brazilian economy during this period. The other has to do with the differences between the deflators used, which may also lead to an overestimation of the mean income growth among those that declare PIT.

The rise in the share of the occupied population that contributes to social security in the PNAD as a measure of the willingness to declare incomes was 2.56 ppy. The idea here is that, as time passed, the IRS became more able to observe incomes. Therefore, the growth of income in the PIT also encompasses the formalization process. The GDP-implicit deflator from the National Accounts grew 1.71 ppy faster than the IPCA. Applying the IPCA to nominal GDP instead of the usual implicit GDP deflator, and taking into account the formalization process, the observed real growth gap fell to 0.54 ppy. In short, we captured an overestimation of financial gains growth, which amounted to an additional impact of 0.35 ppy. The remaining gap of 0.189 ppy to be explained amounts to a small share of 3.86 per cent of the original GDP/PIT income growth gap.

The use of income tax data to adjust for estimates of the income distribution in Brazil assumes that the people who filed their tax returns earn at least what they declared to the IRS. But the argument should not apply to non-taxable income sources. From 2007 to 2015, exempt and non-taxable income gained greater importance, increasing by 9.9 percentage points of the total income filed with the IRS. The income sources that fell the most were all exempt, starting with the exempt income of small and microenterprise owners. This is related to the formalization of the occupied population already mentioned, plus a growing process of transfiguration of workers into legal entities to bypass the costs imposed by Brazilian labour law—an interpretation that is also consistent with the largest income source fall, namely payments from legal entity to declarant.

The fall in exempt retirement income of people 65 years old or above is consistent with reports of a reduction in the number of elderly declarants and their reallocation as dependants of their sons and daughters. From 2007 to 2015, the taxpayer population aged above 70 years fell by 41.6 per cent; more broadly, the population age 41 or above decreased by 15.9 per cent, while according to the PNAD it grew by 30.3 per cent. At the same time, the mean number of dependants per filer up to 40 years of age doubled. This is just the other side of the process that has made elderly tax declarants into their respective sons' or daughters' tax dependants. Note that both these movements are at odds with demographic trends. But why?

Until 2008, each adult had to complete a PIT form in order to have a valid Social Security number (CPF) in Brazil. This was a way to try to control income tax evasion in the country. After 2008, this obligation was dropped, which created an extra incentive to move to dependant status in the PIT records if the dependant is below a given threshold. Furthermore, Brazilian income tax legislation allows the individual to list as dependants both their parents (and the parents of their parents and so on) and incorporate their social security benefits and pensions up to a threshold as exempt income and to discount their health expenses. Hence, if a couple opts for a joint declaration, it can list as dependants all their living parents, grandparents, great-grand parents, and so on. The combination of these institutional and cumulative demographic changes created additional incentives for younger people to incorporate their parents' incomes in their PIT declarations. This would explain the marked rise in exempt income after 2008 and its impact on PIT income growth.

Overall, the paper's main message is that the combination of household survey data and PIT records as opposed to plain survey estimates yields higher levels of Brazilian inequality during the 2007 to 2015 period, as well as different trends. It also yields higher mean incomes in both levels and trends. While social welfare is unequivocally higher in level, in most cases there is also a faster social welfare growth trend associated with this data-combination exercise.

Finally, we also looked at changes in demographics such as the age distribution of individuals who declare PIT and their number of dependants. We showed that some of the changes observed go against the changes in the population profile observed using household surveys, suggesting that there may be incentive effects affecting the share of the population that declares PIT and their respective income levels. This reveals the risk of inferring the trend of Brazilian inequality from PIT tabulations alone and taking their indications at face value.

References

Afonso, R. (2017). 'Transfigurações'. In: J.R. Afonso, M.R. Lukic, R.O. Orair, and F.G. Silveira (eds), *Tributação e desigualdade*. Belo Horizonte: Letramento, Casa do Direito, FGV Direito Rio, pp. 625–70.

Atkinson, A.B. (2015). *Desigualdade: o que pode ser feito?* São Paulo: LeYa.

Barros, K.N.M. (2016). 'Sistema de Contas Nacionais: alguns aspectos sobre a compilação de dados de renda'. 3a Conferência Nacional de Produtores e Usuários de Informações Estatísticas, Geográficas e Ambientais (INFOPLAN). Rio de Janeiro: IBGE.

Bastos, E.K.X. (2012). 'Distribuição funcional da renda no Brasil: estimativas anuais e construção de uma série trimestral'. Discussion Paper 1702. Brasília: IPEA.

Corseuil, C.H.L., M.C. Neri, and G. Ulyssea (2014). 'An Exploratory Analysis of the Effects of the Formalization Policy for Individual Micro-Entrepreneurs'. Report for International Policy Centre for Inclusive Growth.

Ferreira, F.H.G., S.P. Firpo, and J. Messina (2017). 'Ageing Poorly?: Accounting for the Decline in Earnings Inequality in Brazil, 1995–2012'. IDB Working Paper 792. Inter-American Development Bank.

Hoffmann, R. (2017). 'A renda e seu imposto: potencial redistributivo, limite de isenção, progressividade e mensuração da renda'. In: J.R. Afonso, M.R. Lukic, R.O. Orair, and F.G. Silveira (eds), *Tributação e desigualdade*. Belo Horizonte: Letramento, Casa do Direito, FGV Direito Rio.

IPEA (2006). 'Sobre a recente queda da desigualdade de renda no Brasil'. In R.P. de Barros, M.N. Foguel, and G. Ulyssea (eds), *Desigualdade de renda no Brasil: uma análise da queda recente*. Brasília: IPEA, pp. 15–85.

IPEA (2013). 'Duas décadas de desigualdade e pobreza no Brasil medidas pela PNAD/IBGE'. *Comunicados do IPEA*, 159. Brasília: IPEA.

Medeiros, M., and P.H.G.F. Souza (2015). 'Top Income Shares and Inequality in Brazil, 1928–2012'. *Revista da Sociedade Brasileira de Sociologia*, 1(1): 119–32.

Medeiros, M., P.H.G.F. Souza, and F.Á. Castro (2015a). 'O Topo da Distribuição de Renda no Brasil: Primeiras Estimativas com Dados Tributários e Comparação com Pesquisas Domiciliares, 2006–2012'. *Dados - Revista de Ciências Sociais*, 1(58): 7–36.

Medeiros, M., P.H.G.F. Souza, and F.Á. Castro (2015b). 'A estabilidade da desigualdade de renda no Brasil, 2006 a 2012: estimativa com dados do imposto de renda e pesquisas domiciliares'. *Ciência & Saúde Coletiva*, 20(4): 971–86.

MF/DATAPREV (2016). *Anuário Estatístico da Previdência Social—AEPS 2015*, 24. Brasília: MF/DATAPREV.

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.

Morgan, M. (2017). 'Extreme and Persistent Inequality: New Evidence for Brazil Combining National Accounts, Surveys and Fiscal Data, 2001–2015'. WIDER Working Paper 2017/12. Helsinki: UNU-WIDER.

Neri, M. (2009). 'Income Policies, Income Distribution, and the Distribution of Opportunities in Brazil.' In L. Brainard (ed.), *Brazil as an Economic Superpower? Understanding Brazil's Changing Role in the Global Economy*. Washington, DC: Brookings Institution Press, pp. 221–69.

Neri, M. (2014). 'As Novas Transformações Brasileiras'. In: J.P. Velloso (Ed.). *Desacorrentando Prometeu - Um Novo Brasil: Brasil das Reformas e das Oportunidades*, 1st edition. Rio de Janeiro: INAE.

Piketty, T. (2014). *O capital no século XXI*. Rio de Janeiro: Intrínseca.

Saboia, J., and J. Hallak Neto (2014). 'A distribuição funcional da renda no Brasil e sua reversão a partir de meados da década de 2000'. In F.S. Earp, E.F. Bastian, and A.M. Modenesi (eds), *Como vai o Brasil? A economia brasileira no terceiro milênio*. Rio de Janeiro: Ímã Editorial, pp. 143–65.

Sen, A. (1974). 'Informational Bases of Alternative Welfare Approaches: Aggregation and Income Distribution'. *Journal of Public Economics*, 3(4/November): 387–403.

Souza, P.H.G.F. (2016). 'A desigualdade vista do topo: a concentração de renda entre os ricos no Brasil, 1926-2013'. Doctoral Thesis. Universidade de Brasília.