TOP INCOMES IMPACTS ON INEQUALITY, GROWTH AND SOCIAL WELFARE: COMBINING SURVEYS AND INCOME TAX DATA IN BRAZIL[1]

Marcelo Neri[2]
Marcos Hecksher[3]

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 to add more realistic top incomes values into 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 measure rises when higher top incomes replace previous estimates from surveys, this same exercise increases by construction, the mean and social welfare levels. Second, while the movement of these combined estimates present a slower inequality fall than pure surveys, mean and social welfare growth trends rose faster. Finally, the paper analyses the nature and causes of a series of measurement error issues. In particular, we are able to reconcile discrepancies between income tax returns, surveys and GDP growth rates.

Key-Words: 1. Top Incomes; 2. Income Inequality; 3. Personal Income Tax Records; 4. Combining data sets; 5. Pareto Interpolation

RESUMO

Este artigo avalia os impactos da combinação entre pesquisas domiciliares e registros administrativos do imposto de renda em termos de crescimento da renda, desigualdade e bem-estar social no Brasil, entre 2007 e 2015. Este exercício guarda a promessa de adicionar às pesquisas domiciliares valores mais realistas do topo da distribuição de renda. Enquanto a literatura acerca do tema focou nos impactos da combinação destas bases de dados sobre a desigualdade de renda, este artigo avalia suas implicações cumulativas sobre bem-estar. Em primeiro lugar, apesar do nível de desigualdade calculado ser maior quando há substituição das rendas mais altas estimadas originalmente nas pesquisas domiciliares, o mesmo exercício de substituição, por construção, também aumenta o nível médio de renda e bem-estar social estimados. Ademais, enquanto o movimento destas estimativas combinadas resultam em uma queda menor de desigualdade do que a observada nas pesquisas domiciliares tradicionais, a tendência de crescimento da renda e do bem-estar social se torna maior. Por fim, analisamos a ocorrência de uma série de erros de medida advindos deste exercício de combinação de bases de dados e as causas associadas. Em particular, reconciliamos as discrepâncias obtidas entre as taxas de crescimento da renda nos dados do imposto de renda, das pesquisas domiciliares e do PIB.

Palavras-chave: 1. *Top Incomes*; 2. Desigualdade de Renda; 3. Dados do Imposto de Renda; 4. Combinação de Bases de Dados; 5. Interpolação de Pareto

Código JEL: C81; D31

Área 12 - Economia Social e Demografia Econômica

---

[1] This paper is part of the Brazilian chapter of the "Inequality in the Giants" project supported by UNU-Wider. We thank the comments provided at the conference 'Income redistribution and the role of tax-benefit systems in Latin America' held in Quito in July 2018.

[2] FGV Social and FGV EPGE.
[3] IPEA and ENCE/IBGE. Also Acknowledges the financial support provided by RPCAP from FGV.

TOP INCOMES IMPACTS ON INEQUALITY, GROWTH AND SOCIAL WELFARE: COMBINING SURVEYS AND INCOME TAX DATA IN BRAZIL

1. INTRODUCTION

Recently, there have been a few studies in Brazil combining data from household surveys with administrative records of personal income taxes (PIT) files, whose main benefit would be to measure more accurately the income of the wealthiest in the population. These studies have applied to the Brazilian case those methods ever more widespread by authors such as Piketty (2014), Atkinson (2015) among others, to estimate income inequality series over the years and even centuries in different countries. Medeiros et all. (2015a; 2015b) have been pioneers in harmonizing these different sources of information in Brazil. More specifically, they have reconciled microdata from the Annual 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 singles out those 10% richest from the rest by replacing PNAD´s adult population highest income by the highest income reported in the IRPF. The hypothesis here is that the latter source has a greater capacity to identify the highest incomes.

This literature emphasizes the impact on income inequality albeit not detailing the 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 it attends the Pareto efficiency criteria, 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-IRPF distribution vis-à-vis the single PNAD distribution. If the portrait of income distribution is considered, the mixed distribution is definitely higher per se. In short, if Brazil followed the income distribution pattern of the mixed PNAD-IRPF 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 in the population. A similar story seems to hold also for comparing income distribution across time. Meaning if inequality presents a slower falling trend, both mean income and social welfare measures increase in the period analysed in these papers. Firstly, we need to analyse the related changes in social welfare, both in levels and 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.

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. Firstly, we include the period after 2012, when a major Brazilian recession began. Secondly, we test different fittings in the distributions and evaluate their implications by using different aggregate measures of 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. Another related problem whose causes have not been properly analysed is that the growth rate of mean income may be overestimated in the PIT hence also in the combined PNAD-PIT.

The paper is thus organized: 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 IRPF 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 for the high rate of growth in PIT income, which may be the least probed issue in the related literature and is behind social welfare trends observed. The fifth section analyses desegregated income sources, in particular the increase of exempt incomes. The sixth section addresses demographic issues and economic incentives that determines the change of the number of PIT declarations by age groups. It is an illustrative example of PIT

tables challenging somewhat established demographic facts. The sixth section brings the main conclusions of the paper.

2. SHORT LITERATURE REVIEW

The 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, precisely the most important ones to determine the inequality level according to usual measures such as the Gini and the Theil-T indexes. Medeiros, Souza and Castro (2015a; b) while analysing filed for 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. The conclusion that inequality did not decrease in this period is reinforced in Souza (2016) thesis that applied different methods to use tax data either isolated or in conjunction with household data to create new Gini index series.

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

The expectation that PNAD particularly underestimates the highest incomes is justified by the patterns of differential non-response to household survey. Firstly, interviewers may have more obstacles in accessing the wealthiest households, which hinders their capacity to capture their income information. Besides, the survey may have serious limitations to correctly measure the wealthiest´s share of income from sources such as rents, interests and profit – as opposed to income from work or social benefits, which tend to be more accurate. Conversely, we have seen the possibility that in income tax return files, revenue from financial investments may be overestimated, as it does not consider the monetary correction, enclosed in the nominal variation that banks report as income.

In any case, income tax return files may be the best tool to obtain an estimate closer to the real income level of the richest, and despite that, they can be inadequate to estimate the variation rates of these incomes through time, which is crucial for the inequality trajectory. According to Medeiros, Souza and Castro (2015a; b), from 2006 until 2012, the income of the wealthiest population drifted away from the mean, contrary to what PNAD indicates. This conclusion can only be legitimate if the underestimation of the wealthiest's income had turned deeper 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 annual 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 in higher taxes paid. Measurement error may also vary through the years. Incentives and the margin of choice to file or not file, each fraction of one´s income may be influenced by legal aspects. Just as in the case of those well-paid workers who acquire the status of legal liability companies for instance, whose profits are then exempt from personal income taxes (AFONSO, 2017). Tax collection and enforcement was also improved. All of these factors can eventually affect share of the wealthiest population´s income to be captured by the Inland Revenue Office, thus influencing its growth rate.

Information that is more detailed may indicate a strong growth in capital income, for instance, overlooked at the PNAD. In principle, the path of the net operational surplus´s share in the functional income distribution, according to the National Accounts, would not validate this hypothesis (BASTOS, 2012;). Nor does the path of interest rates and the distribution of real estate estimated value of both poor and rich, according to the 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. After all, it is not enough to point out a "real path" of income inequality 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 close 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 step by step their closer determinants.

3. COMBINING INFORMATION FROM SURVEYS AND INCOME TAX FILES

3.1 Combination - We have chosen to combine PNAD data with income tax files tabulations according to the methods used by Medeiros, Souza e Castro (2015b). Between the first (2007) and the last (2015) income tax tabulation as disclosed by the Inland Revenue Office on the internet in a standard format as when this chapter was written, the wealthiest population´s personal income was calculated and then combined with income in the PNAD for the remaining population. Afterwards, once we combined the database from PNAD and the Inland Revenue, we analysed changes not only in inequality but also in the mean and in social welfare.

Disclosed in 2017 on the internet[4], the income tax tables here sum up the large figures in taxpayer files in 2008 and 2016 regarding their revenue, respectively along 2007 and 2015. All files, revenues, deduction, taxes, assets and liabilities are disclosed in each table according to different categories, such as sex, city, age group, nature of labour, type of occupation or monthly income in terms of minimum wage (MW). Tables used in this section summarize files according to total revenues from "up to ½ MW" to "more than 160 MWs" (in 2007) or "more than 320 MWs" (in 2015) in monthly terms.[5]

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 until a certain percentage of the population, is correctly filed in the Inland Revenue database. The objective of this estimate is not to point out the "real" value of income, but just to mitigate any underestimation in household surveys.

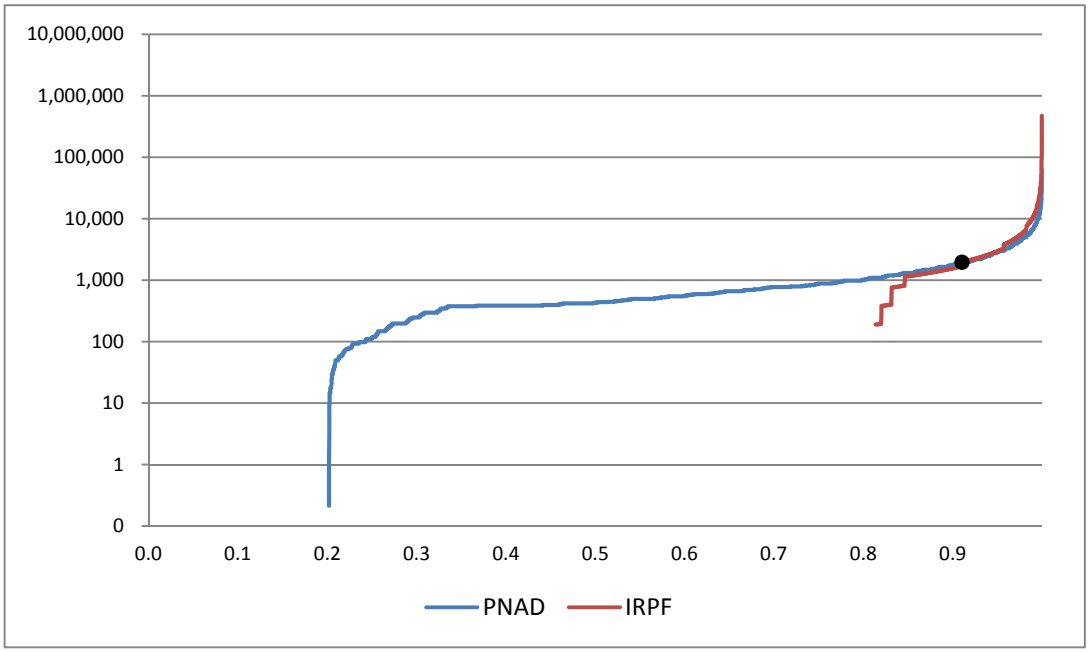
According to Souza (2016), the total population is based on IBGE´s estimate for the number of people living in Brazil on July 1st each year who are 20 years old or older. In line with the mentioned authors and others, we applied a Pareto interpolation on these datasets to estimate the distribution within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

Once the model was 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 PNAD. There are countries where income tax files cover more than 90% of the population. In Brazil, they correspond approximately to a fifth of the adult population, albeit not the richest fifth. Although they also provide fittings into alternative quantiles, Medeiros, Souza e Castro (2015b) affirm that, based on data from different years, income from the tenth richest can be estimated according to the income tax files, and the income from remaining nine tenths according to the PNAD.

[4] Various years available: 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.

[5] Disclosed information increased in detail during the last years, but more can be done if tabulations are also disclosed while integrating data on people and companies. Unidentified longitudinal microdata samples allow to track the same people over the years, respecting confidentiality, allows to estimate different features of individual income processes.

Individual Monthly Income by Population Quantile in 2007 (R$)



Source: PNAD/IBGE; IRPF/RFB

The last graph shows excerpts where the PNAD and IRPF estimates overlap. The chosen fitting points require an increase in the PNAD´s income of the 8.9% richest in 2007 and the 11.4%, 10.0% or 8.9% richest in 2015.

Once these absolute fittings in the distributions are applied[6], in relation to the PNAD, replacing the highest individual incomes that were informed in the PNAD by estimates based on the highest incomes as filed in the Inland Revenue increases the share of total income concentrated by the rich as well as the inequality among the richest. These two changes increase the inequality in the combined PNAD-IRPF database.

The three following tables show that the choice between the different fitting points considered in 2015 barely reflect on the results. While PNAD´s average real income increases 1.7% per year, the average real income of the PNAD-IRPF database increases 2.9% 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 only 0.001 point per year in the PNAD-IRPF 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 2.7% per year in the PNAD, but increases 4.0% per year in the PNAD-IRPF database fitted in the 0.911 quantile. This last rate remains 3.9% when the fitting is done in the 0.886 or 0.900 quantiles, which does not qualitatively alter the analysis.

[6] Thus is called the fitting method adopted by Medeiros, Souza e Castro (2015b). By including periods of high inflation into the analysis, Souza (2016) has chosen another method for estimating the inequality, based on the Gini index´s capacity to split itself into groups that do not overlap.

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%

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

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: PNAD/IBGE; IRPF/RFB and Combined databases.

There was a strong increase in the incomes filed for taxes between 2007 and 2015. This growth in the income tax was more intense than the average growth observed in the PNAD, and in this survey, 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 results thus in an annual growth rate higher for the PNAD-IRPF database (2.9%) than in PNAD´s average (1.7%). However, the average growth remains, even in the PNAD-IRPF database, lower than the growth in the median income (3.0%) 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 to 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 tables 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 (1976) that results from multiplying mean income by the Gini inequality index complement. The welfare level – which in 2007 was 1.02% higher for the combined distribution vis-à-vis single PNAD – increases to 2.41%; thus, causing the

welfare gain between 2007 and 2015 as presented by the combined database to grow 3.2% per year against 3.0% 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 expressed as welfare changes is more propitious in the combined database than in the single PNAD.

Social Welfare (Sen 1976)

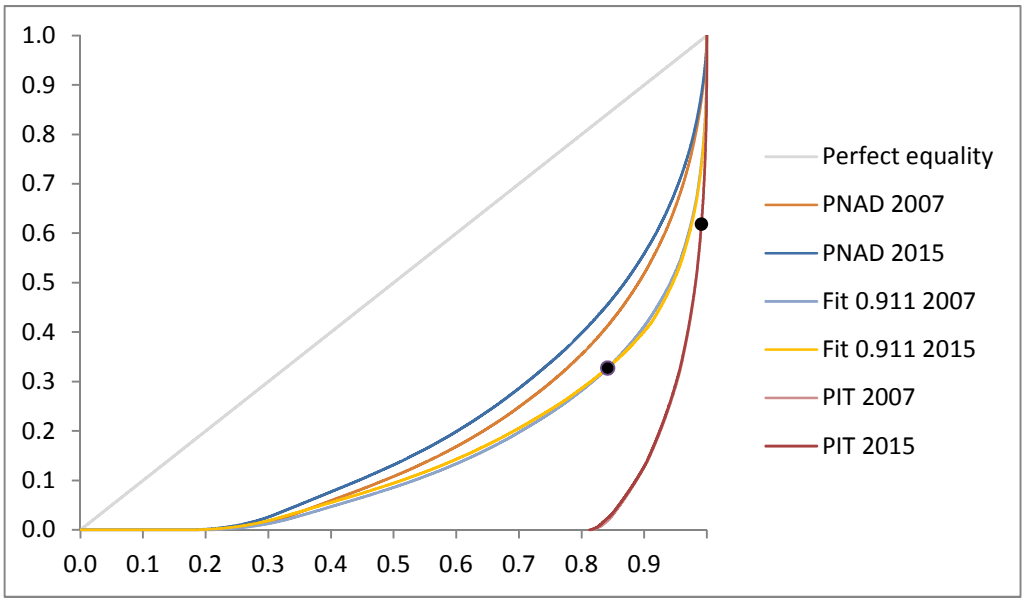
	2007	2015	total var. %	annual var.%
PNAD	500	636	27.2%	3.0%
Fit 0.911	505	651	28.9%	3.2%
Across Bases	1.02%	2.41%	1.7%	0.2%

Source: PNAD/IBGE and combined PNAD-IRPF 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 remains necessarily 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. Besides, variation in the index depends on its sensitivity to each part of the distribution. The Gini index is more sensitive to variations in those quantiles closer to the median, while 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 following Lorenz curves show, in the vertical axis, the percentage of income accumulated by the poorest population until each quantile as 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. PNAD 2007 curve overcomes the PNAD 2015 curve completely, which proves that inequality fell in this period, whatever the index used.

Lorenz Curves for PNAD, IRPF AND PNAD-IRPF databases in 2007 and 2015

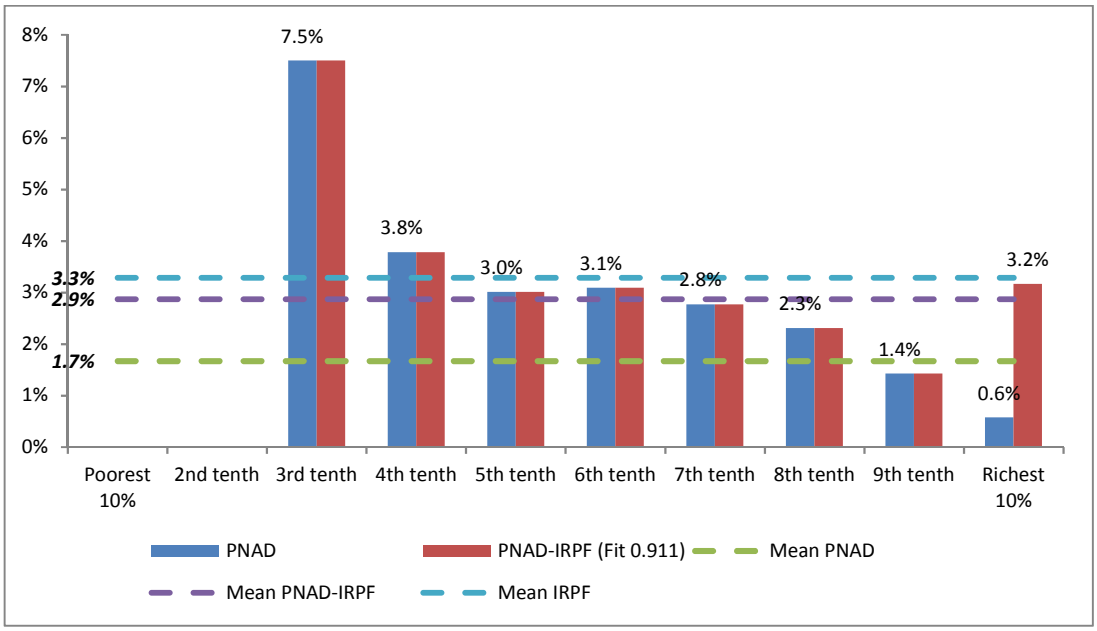


The PNAD-IRPF integration produces Lorenz curves more distant from perfect equality, but the curves for the integrated bases in 2007 and 2015 meet each other. When two Lorenz curves meet, there are indexes indicating a rise in inequality, while other will indicate a decrease. It was exactly what happened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

The graph 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% poorest remained with 32.8% of the total income, while the 15.8% richest remained with 67.2% of total income. Nonetheless, the 2007 curve is higher than 2015´s in all 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 overcomes 2007´s in all 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 its share in the total income? In the PNAD-IRPF integrated database, the seventh, eighth and ninth tenths were the only ones who lost their shares in total income because their growth rates were lower than average (2.9% per year) as the next graph shows.

Real growth rate of income per tenth of the population per year (2007-2015)



Source: PNAD/IBGE; and Combined PNAD-IRPF databases.

It is worth mentioning that the tenth with the lowest growth was the ninth (1.4%) even though at a positive rate. This group comprised people with income close to the average value that 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% of adults had null income in 2007. This percentage drops to 17.7% in 2015. In other words, 2.5% 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% who remained with a null income between 2007 and 2015. The graph reveals that the poorest up until 0.6 quantile of the adult population increased their share of total income. The 10% richest also had a growth rate (3.2%) higher than average (2.9%) but not as high as rates observed in the fourth and fifth tenths (7.5% and 3.8%, 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% 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 PNAD, the income of the 40% poorest among adults aged 20 year-old or more (including those without an income) increased at an average rate of 5.1% per year in real terms, more than the PNAD average (1.7%) and even more than the average in the PNAD-IRPF database (2.9%). The 40% poorest increased their share in total income from 5.9% to 7.7% in PNAD, and from 4.7% to 5.5% in PNAD-IRPF.

The following table 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%, 1%, 0.1% and 0.01%. The maximum value in the PNAD-IRPF database in each year corresponds to the value applied by interpolation to the highest personal income.

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

Income Distribution in PNAD, IRPF and PNAD-IRPF databases in 2007 and 2015

2007 (at 2015 prices)							**Share of total income (%)**	
	Mean		Minimum		Maximum		PNAD	PNAD-IRPF911
	PNAD	IRPF	PNAD	IRPF	PNAD	IRPF		
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							**Share of total income (%)**	
	Mean		Minimum		Maximum		PNAD	PNAD-IRPF911
	PNAD	IRPF	PNAD	IRPF	PNAD	IRPF		
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. 2007-2015 (% a.a.)							**Real var. 2007-2015 (p.p.)**	
	Mean		Minimum		Maximum		PNAD	PNAD-IRPF911
	PNAD	IRPF	PNAD	IRPF	PNAD	IRPF		
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: PNAD/IBGE; and Combined PNAD-IRPF databases.

Is it possible then that PNAD has increasingly missed the income of the richest and hence missed the "miracle" they experienced? Or is it that the income tax became more accurate in measuring the income of a larger group, pointing to a fantastic growth precisely because it identifies better previously unseen incomes?It is not easy to answer this question, not without accessing Inland Revenue microdata or at least tables that show longitudinal movements of taxpayers, dependents 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 Inland Revenue tabulations and the attempts to integrate them with household surveys. The use of tax data disclosed in the last years enables new hypotheses but does not warrant by itself definite conclusions on whether income inequality and income distribution as a whole have varied in Brazil.

4. THE "ECONOMIC MIRACLE" IN THE PERSONAL INCOME TAX TABLES

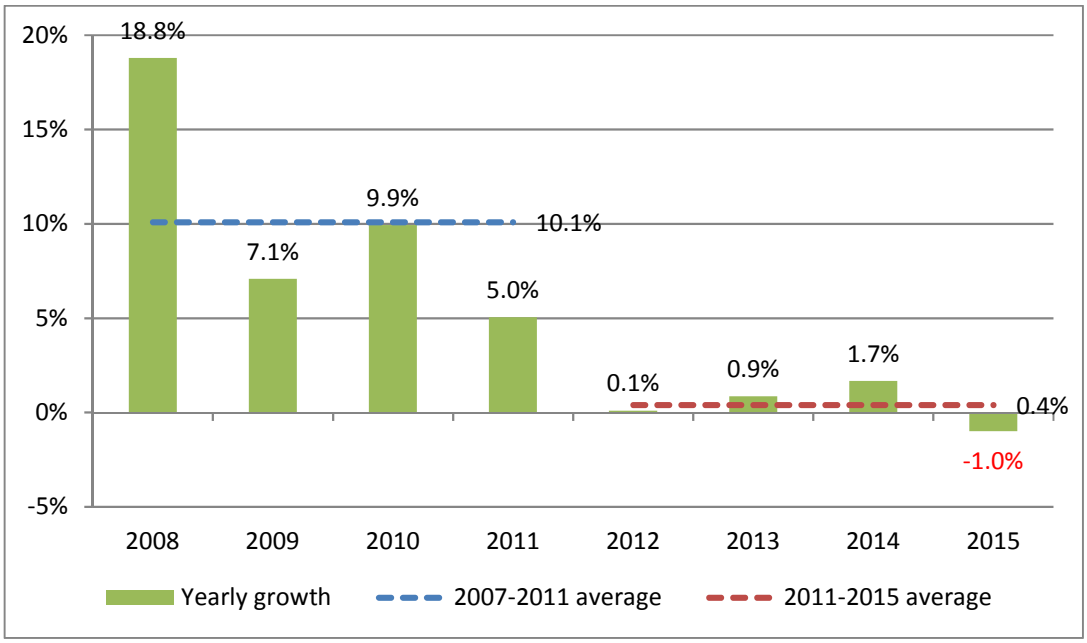
If one wants to understand the combined PIT-PNAD distribution levels and changes related facts and sources, we should also analyze Personal Income Tax (PIT) data in isolation. PIT figures as released by the Brazilian Inland Revenue (RFB), available in standard format for the 2007-2015 period. PIT mean income growth for the 2007-2015 period presents astonishing differences with respect to GDP growth: 4.97% per year against 1.23%. That is a 304% faster growth rate, or a gap of 3.74 percentage points per year (ppy). If we believe in PIT income data at face value, this difference yields non negligible impacts on social welfare growth in the period. But perhaps, more relevant, 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 social welfare and inequality changes since they all come from the same data set.

4.1 Mean Income Growth – For the same reason, it is worth breaking the analysis into different periods. Calls attention the impressive real growth rate of the mean income per declarant between 2007 and 2011, i.e. 10.1% per year when applying the IPCA (Wide Consumer Price Index) as deflator. During the same period, the Brazilian Geographical and Statistical Institute (IBGE), Brazilian reveals that per capita GDP grew 3% per year.

Between 1967 and 1973, Brazil experienced the so-called "economic miracle" for its high annual growth rates. In that period, the per capita GDP grew 8.3% per year on average, as the graph below shows. Since the beginning of last century, per capita GDP increased more than 10% only in five specific years: 1901, 1920, 1928, 1936 and 1973, last year of the so-called "miracle".

In terms of income tax, the average income per file increased 18.8% in 2008, 7.1% in 2009, 9.9% in 2010 and 5.0% in 2011, according to the graph below. In the next four years, growth in the Income Tax was milder, about 0.4% per year on average, closing the series with a decrease of 1.0% in 2015 when the per capita GDP dropped 4.6%.

Real Growth Rates of Mean Income per income tax file (% a.a.)



Source: RFB. Deflation by IPCA/IBGE.

PIT mean income growth for the 2007-2015 period presents astonishing differences with respect to GDP growth: 4.97% per year against 1.23%. That is a 304% faster growth rate, or a gap of 3.74 percentage points per year (ppy).

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 millions in 2007 to 27.52 millions in 2015 a growth rate of 1.1 ppy in the period rising the gap to be explained to 4.88 ppy.

4.2 Propensity to Declare – One line of reasoning has to do with increasing formalization of the Brazilian economy during the 2007 to 2015 period, which may lead to an overestimation of the mean income 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 practise. 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 of the share of the occupied population that contributes to social security in PNAD. It rose from 50.3% in 2007 to 61.52% in 2015, an annual

growth rate of 2.56% that is explains more than half of the gap 4.88 ppy leaving still 2.26 ppy to be explained[7].

4.3 Deflators – Another subsequent 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 wider CPI (IPCA)[8]. 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 reconciling almost all differences between GDP and standard PNAD income growth differences in the 2003 to 2013 period. These annual differences turns out quite similar to the 2007 to 2015 period. Applying to nominal GDP the wide CPI (IPCA) instead the usual implicit GDP deflator, the observed real growth gap would fall to one half from 2.26 percentage points per year (ppy) to 0.54 ppy.

Looking into the data by type of income implemented immediately below allows to advance in the income growth gap puzzle solution. Making a long story short we captured an overestimation of financial gains growth that amounted to an additional impact of 0.35 ppy. This remaining gap of 0.189 ppy to be explained amounts to a small share of 3.86% of the original GDP/PIT income growth gap.

5. INSPECTING INCOME TYPES: THE GROWTH OF EXEMPT SOURCES

5.1 Overview - The use of income tax data to adjust for estimates about the income distribution in Brazil assumes that the people who filed their tax returns exist and earn at least what it was declared to the Inland Revenue RFB (Morgan, 2017, p. 3). It is true that it is not incentive compatible in general to declare higher taxable income than it is observed in fact 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 challenges the initial idea that lead to the replacement of PNAD data for PIT records on the top incomes range.

From 2007 to 2015, exempt and non-taxable income gained greater importance, increasing from 21.4% to 31.3% (+9.9 percentage points) of the total income filed in the RFB, while taxable revenues reduced their share from 70.7% to 58.9% (-11.8 p.p.) and income subject to exclusive tax increased their share from 7.9% to 9.8% (+1.9 pp). One has to infer about the reasons behind this dramatic increase on the share of non-taxable income.

Making a zoom 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 less – was the amount received from employers/companies by the taxpayer, a taxable revenue and the largest of them all. The share of this source dropped 7.6 pp from 60.8% to 53.2%. The income sources that grew the most were all exempt, notably three: small and microenterprises (+2.4pp), the exempt share of retirement income of +65 years old people (+1.3 p p) and savings account income and mortgage debts (+1.2 pp).

5.2 Small Business Formalization and Transfiguration - The growth of exempt small business owner income may relate to increase in their profits, but the main causes 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, to facilitate the hiring of workers as companies 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 capitalists to the RFB. This result is consistent with the greatest fall of the 48 income sources

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

[8] In the 2007 to 2015 period he GDP implicit deflator at market prices grew 7,96% against 6,14% of the CPI (IPCA). One may argue that the later is better suited to measure social welfare changes.

pointed above which is a taxable income item: payments from legal entity to declarant which fell 7.6 pp in the 2007 to 2015 period.

The number of files pertaining to partners or owners of microenterprises who perceived income from profit and dividends increased their share in the total number of files in 4.5 pp from 4.3% to 8.8% between 2007 and 2015, or 122,8% which corresponds to 10,5% per year. In the same period, The share of employers and self-employed according to PNAD grew only 0,44% per year[9]. In this interim greater formalization of businesses 12,4% per year rise according to PNAD 2007-2015 as a result of the new 2009 MEI (individual microentrepreneur) Law which was quite effective and also led to the transfiguration of employees into self-employment (Courseil and Neri 2014).

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. It is a result of a combined increase in the concentration of benefits with lower values of up to 2 minimum wages, and the growing number of older people receiving these benefits (MF/DATAPREV, 2016).

This phenomenon interacts with the one scrutinized in chapter 6 that reports a reduction in the number of elderly declarants and their realocation as dependents of their offspring (or even grandchildren). The institutional change that apparently triggered this movement is also presented there. For now it suffices to note that there was also a conversion of older taxpayers into dependants, that at least partially explain the drop of 3.1 million in the number of tax return files among those more than 41 years old between 2007 and 2015. From 2007 to 2015, taxpayers population aged 41 or + decreased 15.9% in the period, from 19.7 to 16.6 million. While in PNAD it grew 30,3%. At the same time the average number of dependents per person with up to 40 years of age has doubled from 1 to 2 dependents for each 3 files within this group. That is a growing share of people aged 41 years old and above were registered as economic dependents. This is just the other side of the coin. This would help explain the intriguing decrease in the number of +41 years old filing their tax returns, in the opposite direction of demography.

It is also true that there was strong increase in the share of people with up to 40 years old who file their income tax returns. It is true that there was also a relative gain in the income of younger workers, as caused by the reduction in the return on the experience. According to Ferreira, Firpo e Messina (2017), this was the main determinant of the decrease in the labour earnings inequality between 1995 and 2012. These authors raise two possible explanations for this phenomenon. The first one would be an age-biased technical change, following a greater demand for younger workers who are more apt to produce with new technologies. The second would be a shift in the Brazilian production towards those sectors benefiting 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 is fraught with an overestimation of the financial investments´ income in the income tax files (Hoffmann, 2017), which is aggravated between 2007 and 2015. We saw increases in specific groups non-taxable incomes that will be detailed now. The last point raised above is consistent with the overestimation on the income processed in the RFB tables because it uses information on nominal instead of real interest incomes (Hoffmann, 2017, p. 385-386). 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, informed to the RFB 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, p. 386) may explain the positive correlation that he points out between the Brazilian annual inflation rate and the inequality indexes reported yearly by Medeiros and Souza (2016) and Milá (2015) based on income tax

[9] The share of employers according to PNAD remained stable at 3.7% while the share of self-employed grew 1.7 pp from 21.2% to 22.9%.

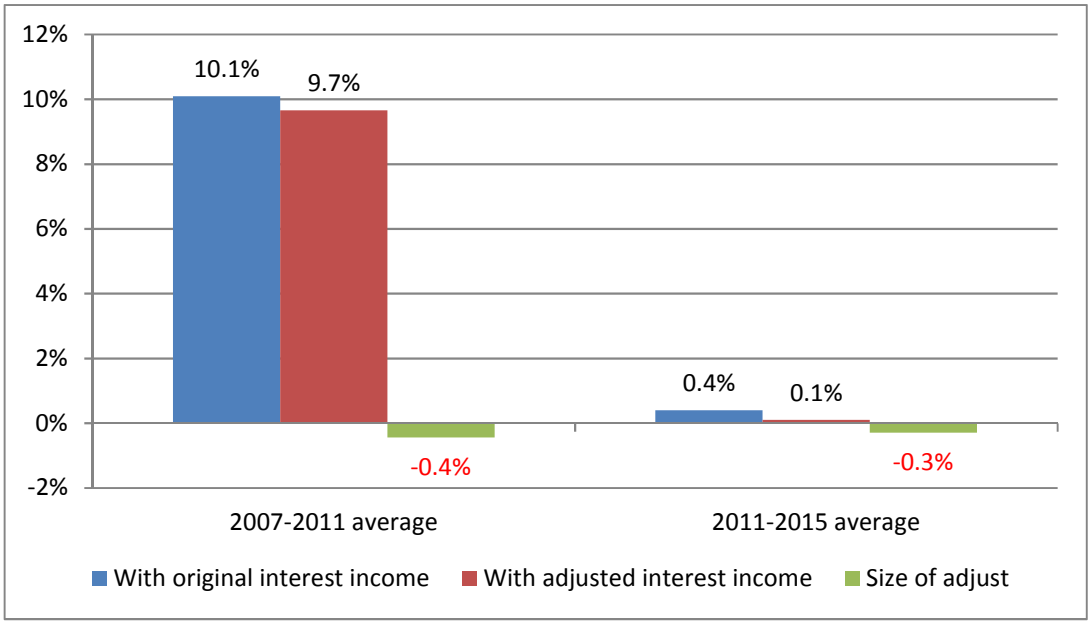
return files data. These correlations overcome 0.8 and are statistically significant at 1%, even though there have been only eight observations in the series analysed by the author.

To estimate the real income from interest equivalent to the nominal income presented in the RFB tables, it is not enough to deflate the filed values. Besides inflation, it is necessary to know what is the amount corresponding to the flow of interests or the profitability rate of each investment. RFB tables show total values, on December 31st 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).

Making a long story short, if the series on total income were adjusted to exclude these estimates of monetary indexation of savings accounts and financial investments, the average annual growth rate of total income per files would be attenuated in 0.4 p.p. in the 2007-2011 period and in 0.3 p.p. in the 2011-2015 period. The following graph shows that, even after this adjustment, the 2007-2011 period still remains with a 9.7% rate per year of real growth of average income per file. The "economic miracle" in the income tax tables remains, even if the "monetary illusion" of these two financial assets are excluded.

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



Source: author´s own estimates based on RFB data

6. TAXPAYER´S PROFILE CHALLENGES DEMOGRAPHIC PATTERNS

The rationale of analysis of the Personal Income Taxes (PIT) has to incorporate the existence of incentives and hence selectivity biases that makes conclusions taken based on that, or on the combination of PIT records with household surveys more complex.

6.1 Number of files - The simple evolution in the number of files processed by the RFB is also surprising. 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 1.2% from 25.2 million to 24.9 million. From 2011 to 2015, in its turn, the number of files increased 10.5% reaching 27.5 million, despite the economic slowdown. For the whole period, between 2007 and 2015, the number of files grew 1.1% annually on average.

The limit of income tax exemption in current values increased in a linear fashion, at 4.5% per year. The IPCA inflation rate overcame this nominal adjustment each year, reducing the real value of the exemption cap year after year. The minimum wage increased in real terms (above the inflation) hence the exemption cap drops further in terms of minimum wages. In 2007, monthly income of up to 3.37 minimum wages were exempt from the income tax. In 2015, just monthly income below 2.38 minimum wages were exempt.

If exemption rates 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. Besides, the following graph shows that the number of files for total income of up to 3 minimum wages did not increase between 2007 and 2015, although this group benefited from the 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 minimum wages increased from 18.6 to 21.1 million.

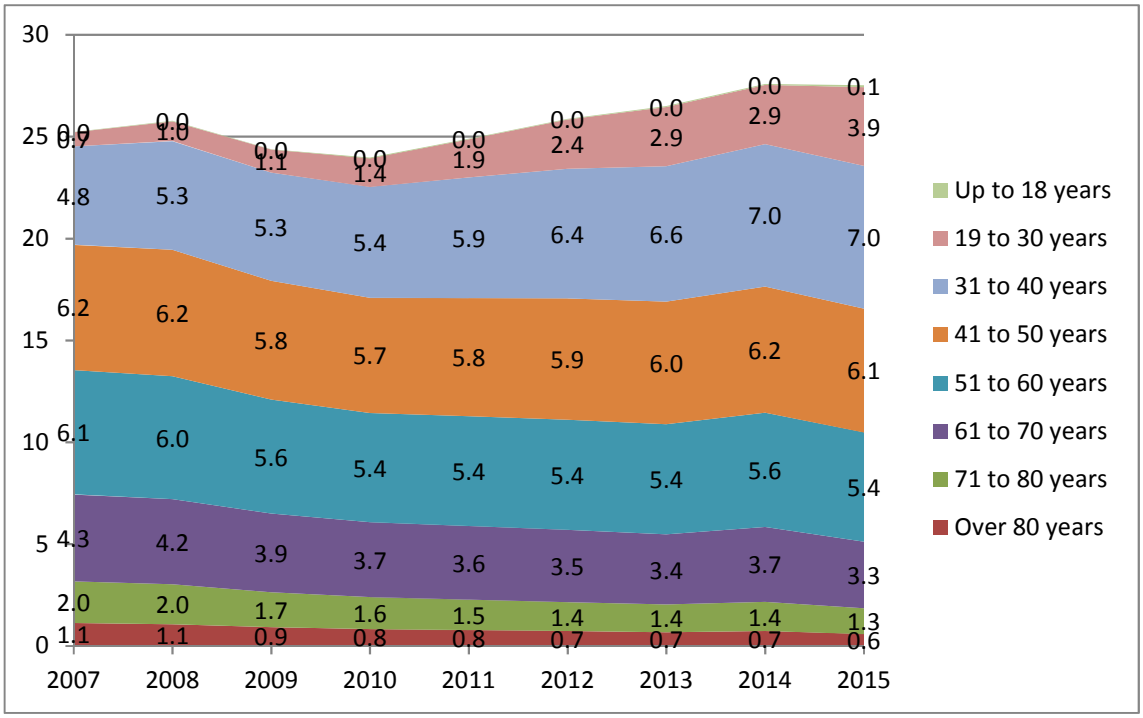
The number of files concerning 2 to 3 minimum wages grew the most, both in relative terms (+83.2%) as well as absolute terms (+1.5 million files). Nonetheless, the number of files for total income below 2 minimum wages dropped in 1.7 million, with decreases of 51.4% for the group earning up to ½ minimum wage, or 24.0% for those earning between ½ and 1 minimum wage, and 11.5% from 1 to 2 minimum wages. At the same time, in the other extreme, tax return files for incomes above 160 minimum wages increased 13.2%, from 66.6 to 75.4 thousand.

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 through 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, the income taxpayers are getting younger. From 2007 to 2015, taxpayers aged 41 or more[10] decreased not only their relative share in the total tax return files, but also its absolute number that decreased 15.9% in the period, from 19.7 to 16.6 million. The IBGE estimates that the 41+ year-old group increased 25.1% in the same period in the total population, from 55.3 million to 69.2 million people, while the 15-40 year-old group only grew 0.7%. Nevertheless the total number of processed tax return files only increased (+9.1%) between 2007 and 2015, from 25.2 million to 27.5 million. As the next graph shows, the mentioned decrease in the number of 41+ years old filing tax returns was outweighed by the 98.0% boom in the number of taxpayers up to 40 years old, from 5.5 million to 11.0 million.

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

Number of Income Tax Files per age group (millions)



Source: IRPF/RFB.

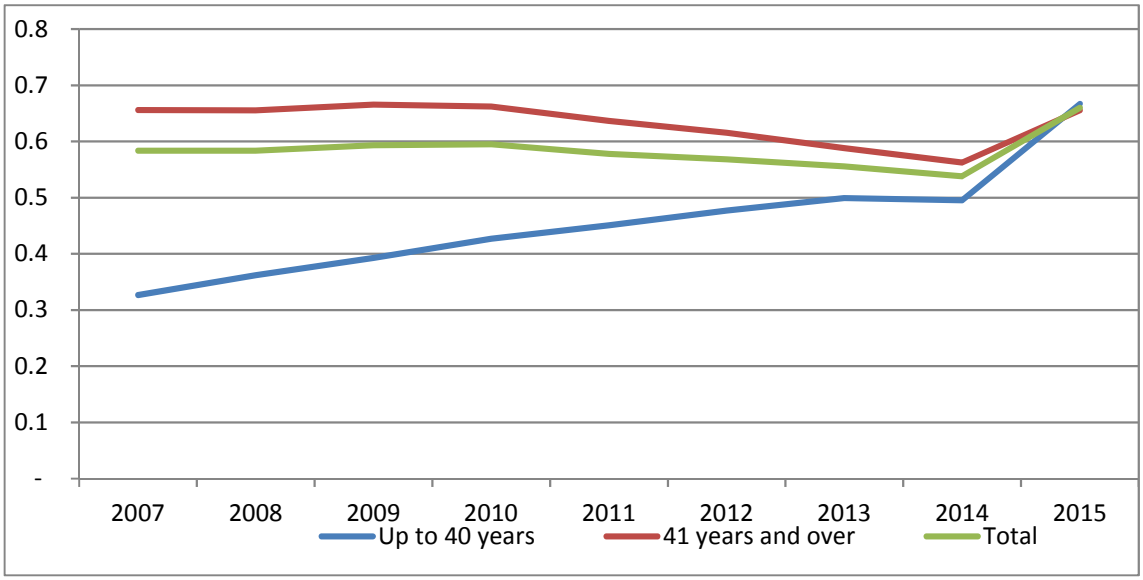
6.3 Dependents - The income pertaining to each file does not reflect the income of a single person, but the income of the person who filed the tax return plus income from his/her dependents. Changes in the number of dependents per file and their allocation into other groups may bias the analyses, and there is enough evidence to suggest that there have been crucial changes in the period of interest. Tables from RFB do not display the number of dependents for each group of files, but a proxy for this number is the total amount of deductions per dependents in each group, which are available in the tables, divided by the annual amount of deduction by each dependent, whose historical series are available at the RFB website.[11]

According to this estimate, the following graph shows that the estimate average number of dependents per person with up to 40 years of age has doubled from 1 to 2 dependents for each 3 files within this group, matching the average for those over 41 years old. That is a growing share of people aged 41 years old and above were registered as economic dependents. This would help explain the intriguing decrease in the number of +41 years old filing their tax returns, in the opposite direction of demography, but does this hypothesis make sense?

According to the National Household Sample Survey (PNAD), the average number of inhabitants in each household is decreasing in Brazil from 3.36 in 2007 to 3 in 2015. Nevertheless, dependants of a tax return filing person do not need to live in the same household as him/her.

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

Estimated average number of dependents according to the age of main taxpayer



Source: IRPF/RFB.

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 do not reach a certain cap, which in 2017 was R$ 1,903.98 monthly. Taxpayers who may file their taxes together or separately may chose the most advantageous option. The joint tax return file is a better option when the largest share of the +41-year-old person´s income is exempt from tax, thus preventing the joint file to lead to higher tax rates. Moreover, it could also enable a deduction in the total due tax as the dependant may have medical or other expenses, which would otherwise be innocuous in the case of exempt income.

Until 2008 each declarant had to fill a PIT form to have a valid Social Security number (CPF) in Brazil. This was a way to try to control income tax evasion in the country. Since 2008, this obligation was abandoned which created an extra incentive to move from declarant to dependent status in the PIT records if it is in the proper income range[12]. Furthermore, Brazilian income tax legislation allows the individual to put as their depends all their parents (and parents of their parents & so on). So a couple if opts for a joint declaration can put both their living parents, grand parents, and so on.

Our main hypothesis here is that the movement of the elderly from declarant to dependent was influenced by the imposition of this new norm. Financial incentives and perhaps, more importantly, reduction of transaction costs could explain the process of piling more income sources within the same PIT declaration. This line of reasoning has the potential to explain few changes that occurred between 2007 and 2008 (as well as lagged effects of this change that might have lasted until 2011 – (between brackets the yearly growth of the 2007-2011 period), namely: i) the sharp rise of mean declared real income of 18,8% (10,1%); ii) the relative fall of the number of elderly declarations -1.23% (-3.5%) with respect to formalized population; iii) the increase in the number of dependents of people below 41 years of age 9,1% (36,4%).

7. CONCLUSIONS

Recent studies combining data from household surveys with administrative records of personal income taxes hold the promise to measure more accurately top incomes. This literature emphasizes the rising income inequality impact of this data-combination exercise, albeit not detailing the effect in mean income

[12] In July 25th, 2008 there was a new norm by the IRS (Instrução Normativa no 864) that abolished the need to present income yearly Exempt Declaration (Declaração Annual de Isento). In 2007 this exemption ceiling was BRL 15,7 thousand in taxable income or BRL 40 thousand for total exempt or other income sources. In 2007, there was 24 million declarations, , 66 million exempt declarations which corresponded to around 90 million active social security numbers (CPFs). On top of that there were 48 million suspended or waiting for regularization - social security numbers.

and in 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 by construction, 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 pioneer work by Medeiros, Souza and Castro (2015a; b) applied this type of approach to Brazil between 2006 and 2012. They have shown indeed that that inequality as measured by the Gini index is 11% higher in the final year and decreased 2 percentage points less in the period, than the pure traditional PNAD estimates. Conversely, our calculations on their published statistics informs that the Brazilian income would also be 35% higher and would have grown 13 percentage points more between 2006 and 2012. If we were to solve this tie in qualitative results with impacts on social welfare by using Amartya Sen (1976) specification, the level of the combination of mean income with its inequality measured by the Gini index on a synthetic index of social welfare would be 9.62% higher and would grow 11 percentage points more in the analysed period. However, more generally, the growth incidence curve comparing the two extremes years is always in the positive quadrant showing an unequivocally increase in social welfare - in fact a Pareto improvement - during this period.

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

We considered the population aged 20 or more and their individual income from all sources. We perform a series of robustness tests on the compatibility of both surveys populations and explore different income quantile fittings. The analysis described here uses the same fitting point in both years (0.911 quantile).We refer again to the social welfare measurement criteria proposed by Sen (1976) that results from multiplying mean income by the Gini inequality index complement. The welfare level in 2015 was 2.36% higher for the combined distribution vis-à-vis single PNAD. Although the Gini was 18.6% higher, the mean income was also 38.5% higher in the same year. The welfare gain between 2007 and 2015 in the combined database is 3.2% per year. The same number in pure PNAD was 3.0% 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 – the 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 changes is more propitious in the combined database than in the single PNAD.

Among all differences between PIT and PNAD the faster income growth trends of the former is the main driving behind the results pointed above. This is 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 the gap to be explained to 4.88 ppy.

To explain this gap, we combined initially two ingredients to explain this puzzle. The first one 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 of the share of the occupied population that contributes to social security in PNAD as a measure of the willingness to declare incomes rose 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 encompasses also the

formalization process. The GDP implicit deflator from the National Accounts grew 1.71 ppy faster than the wider CPI (IPCA). Applying to nominal GDP the wide CPI (IPCA) instead the usual implicit GDP deflator, and taking into account the formalization process the observed real growth gap would fall to 0.54 ppy. Making a long story short we captured an overestimation of financial gains growth that amounted to an additional impact of 0.35 ppy. This remaining gap of 0.189 ppy to be explained amounts to a small share of 3.86% of the original GDP/PIT income growth gap.

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

The fall of exempt retirement income of people 65 years old or above is consistent with reports a reduction in the number of elderly declarants and their realocation as dependents of their sons and daughters. From 2007 to 2015, taxpayers population aged above 70 years fell 41,6%. Or more broadly the population age 41 or above decreased 15.9% while in PNAD it grew 30,3%. At the same time the average number of dependents per person with up to 40 years of age has doubled. This is just the other side of the process that made elderly tax declarants into their respective sons or daughters tax dependents. Note that both these movements are at odds with demographic trends. But why?

Until 2008 each declarant had to fill a PIT form. This was a way to try to control income tax evasion in the country. After 2008, the obligation of filling PIT form to have a valid Social Security number (CPF) in Brazil was abandoned which created an extra incentive to move from declarant to dependent status in the PIT records if it is in the proper income range[13]. Furthermore, Brazilian income tax legislation allows the individual to put as their depends all their parents (and parents of their parents & so on). So a couple if opts for a joint declaration can put as dependents both their living parents, grand parents, great grand parents and so on.

Overall, the paper main message is that if the combination between surveys and PIT records compared to plain survey estimates yield higher levels and trends in Brazilian inequality measures during the 2007 to 2015 period. It also does yield higher mean incomes in both levels and trends. While social welfare is unequivocally higher in levels, in most cases there is also an increasing trend in social welfare growth rates associated with this data sources combination exercise.

Finally, we have also looked at changes in demographics such as age distribution of individuals that declared PIT and their number of dependents. We showed that some of the changes observed goes 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. There is a risky involved in the inferences on the trend of Brazilian inequality using PIT available tabulations.

REFERENCES

AFONSO, R. Transfigurações. In: AFONSO, J. R.; LUKIC, M. R.; ORAIR, R. O; SILVEIRA, F. G. *Tributação e desigualdade*. Belo Horizonte: Letramento; Casa do Direito; FGV Direito Rio, 2017, p. 625-

[13] In July 25th, 2008 there was a new norm by the IRS (Instrução Normativa no 864) that abolished the need to present income yearly Exempt Declaration (Declaração Annual de Isento). In 2007 this exemption ceiling was BRL 15,7 thousand in taxable income or BRL 40 thousand for total exempt or other income sources. In 2007, there was 24 million declarations, , 66 million exempt declarations which corresponded to around 90 million active social security numbers (CPFs). On top of that there were 48 million suspended or waiting for regularization - social security numbers.

670.

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

BARROS, R. P.; CURY, S.; ULYSSEA, G. A desigualdade de renda no Brasil encontra-se subestimada? Uma análise comparativa usando Pnad, POF e Contas Nacionais. *In*: BARROS, R. P. DE; FOGUEL, M. N.; ULYSSEA, G. (Eds.). *Desigualdade de renda no Brasil: uma análise da queda recente.* Brasília: Ipea, 2006. p. 237–273.

CORSEUIL, NERI, M. C.; C. H. L. ; ULYSSEA, Gabriel . An Exploratory Analysis of the Effects of the Formalization Policy for Individual Micro-entrepreneurs. International Policy Centre for Inclusive Growth (IPC - IG) UNDP, v. 129, p. 01-23, 2014.

FERREIRA, F. H. G.; FIRPO, S. P.; MESSINA, J. *Ageing poorly?*: accounting for the decline in earnings inequality in Brazil, 1995-2012. Inter-American Development Bank, mar. 2017. (IDB Working Paper Series, n. 792).

HOFFMANN, R. A renda e seu imposto: potencial redistributivo, limite de isenção, progressividade e mensuração da renda. In: AFONSO, J. R.; LUKIC, M. R.; ORAIR, R. O; SILVEIRA, F. G. *Tributação e desigualdade*. Belo Horizonte: Letramento; Casa do

IPEA. Sobre a recente queda da desigualdade de renda no Brasil. *In*: BARROS, R. P. DE; FOGUEL, M. N.; ULYSSEA, G. (Eds.). . *Desigualdade de renda no Brasil: uma análise da queda recente.* Brasília: Ipea, 2006. p. 15-85.

MEDEIROS, M.; SOUZA, P. H. G. F.; CASTRO, F. Á. 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,* v. 1, n. 58, p. 7-36, 2015a.

_____. A estabilidade da desigualdade de renda no Brasil, 2006 a 2012: estimativa com dados do imposto de renda e pesquisas domiciliares. *Ciência & Saúde Coletiva*, v. 20, n. 4, p. 971-986, 2015b.

MF/DATAPREV. *Anuário Estatístico da Previdência Social* – AEPS 2015, v. 24. Brasília: MF/DATAPREV, 2016.

MILÁ, M. M. *Income concentration in a context of late development*: an investigation of top incomes in Brazil using tax records, 1933-2013. Master Dissertation Paris School of Economics, 2015.

MORGAN, M. *Extreme and Persistent Inequality*: New Evidence for Brazil Combining National Accounts, Surveys and Fiscal Data, 2001-2015. WID, Aug. 2017. (WID Working Paper Series, n. 2017/12).

NERI, M. . As Novas Transformações Brasileiras. In: João Paulo dos Reis Velloso. (Org.). Desacorrentando Prometeu - Um Novo Brasil: Brasil das Reformas e das Oportunidades. 1ed.Rio de Janeiro: INAE, 2014, v. 1, p. 49-92.

PIKETTY, T. *O capital no século XXI.* Rio de Janeiro: Intrínseca, 2014.

SOUZA, P. H. G. F. *A desigualdade vista do topo: a concentração de renda entre os* ricos no Brasil, 1926-2013. Tese de Doutorado. UnB, 2016.