

EARNINGS INEQUALITY IN THE BRAZILIAN FORMAL SECTOR: THE ROLE OF FIRMS, EDUCATION AND TOP INCOMES 1994-2015¹

Marcelo Neri²
Cecília Machado³
Valdemar Pinho Neto⁴

ABSTRACT

This paper documents the evolution and the main close determinants of earnings inequality in the Brazilian formal sector from 1994 to 2015 using matched employer-employee administrative records. In 2015, schooling explained 33% of overall inequality. Firms' specific effects explain 65% of total inequality. Firms are also central to explain the marked inequality fall observed. This downward trend parallels the one seen in household surveys. However, the distributive decompression goes only until the 90th percentile, which is in line with Personal Income Tax based evidence. Specific groups' analysis shows that the share of inequality explained by top 1% and 0.1% incomes rose 43% and 91%, respectively.

Keywords: 1. Earnings Inequality; 2. Linked Employer-Employee Data, 3. Firm and Worker Heterogeneity; 4. Brazil.

RESUMO

Este artigo utiliza registros administrativos de empresas para analisar a evolução e os principais determinantes próximos da desigualdade de renda no setor formal brasileiro entre 1994 e 2015. Em 2015, a escolaridade explicou 33% da desigualdade total. Efeitos-fixos das empresas explicaram 65% da desigualdade total. Empresas também tem papel nevrálgico entre os elementos que explicaram a queda marcada de desigualdade observada. Esta tendência negativa guarda semelhanças com a observada nas pesquisas domiciliares. Porém, a descompressão distributiva vai apenas até o percentil 90, fato que está em linha com as evidências obtidas através dos dados do Imposto de Renda. Análises de grupos específicos revelam que a parcela da desigualdade explicada pelo top 1% e 0,1% mais ricos aumentou 43% e 91%, respectivamente.

Palavras-chave: 1. Desigualdade de Renda; 2. Dados Empregador-Empregados Combinados; 3. Heterogeneidade entre Firms e Trabalhadores; 4. Brasil.

Códigos JEL: D31; I24; J31

Área 13: Economia do Trabalho

¹ This note is part of the Brazilian chapter of the “Inequality in the Giants” project supported by UNU-Wider.

² FGV Social and FGV EPGE, marcelo.neri@fgv.br.

³ FGV EPGE.

⁴ FGV EPGE. Acknowledges the financial support provided by RPCAP from FGV.

EARNINGS INEQUALITY IN THE BRAZILIAN FORMAL SECTOR: THE ROLE OF FIRMS, EDUCATION AND TOP INCOMES 1994-2015

1. Introduction

The vast majority of the empirical literature in developing countries on income distribution is based on household surveys. Brazil established this tradition during the early 1970s just after the second set of the Demographic Census income data was released (Fishlow 1972; Langoni 1973; Bacha and Taylor 1978). Recently, a series of papers have documented inequality based on Personal Income Tax (PIT) records (Medeiros et al. 2015a; Souza 2016; Medeiros et al. 2015b). However, establishment-level administrative records are also available in Brazil, but those have rarely been used in studies of income inequality. RAIS (*Registro Anual de Informações Sociais*) is a matched employer-employee data that gathers around 30 million observations on workers per year over the last two decades by the Brazilian Labor Ministry. RAIS depicts formal employment dynamics and wage differentials and is a powerful tool that may complement the evidence presented by other data sources (Alvarez et al., 2017; Machado et al., 2017).

This paper describes the evolution and the main determinants of earnings inequality in the Brazilian formal sector from 1994 to 2015 using RAIS. We plot growth incidence curves and Lorenz curves over the period of analysis, and calculate the main inequality indexes used in the literature such as earnings ratios across different percentiles in the individual earnings distribution, the Gini index and the Theil indexes. We discuss the role of wages, employment, missing values among other measurement issues. We also compare these results using RAIS with broader household surveys. Second, we use standard inequality decompositions based information theory to understand the main determinants of formal earnings dispersion. This includes workers' characteristics (such as gender, race, age, education and spatial location) and firms' characteristics (sector of activity, firm size, legal nature, etc.). Besides applying between and within groups decomposition for Theil T and Theil L indexes (Theil 1967), we use J-Divergence measures to disentangle the role played by specific categories of different variables (Jeffreys 1946; Rohde 2016; Hecksher et al. 2017).

We find an overall fall in inequality after 1994. Moreover, schooling was responsible for explaining 32.8% of labor income inequality in 2015 and 24.1% in 1994, considering the Theil-T index. The explanatory power of firm-specific was around 65% for the entire series analyzed (1994-2015), suggesting that differences between firms explain the largest share of inequality in the Brazilian formal labor market. These results agree with Alvarez et al. (2017), who found that firms played an important role in explaining inequality level and also the decrease in earnings inequality in Brazil. It is important to note that the between-firm component also seems to drive the overall inequality in developed countries such as the U.S. (Song et al., 2015) and Germany (Card et al. 2015).

Changes in earnings distribution in the formal sector share some of the trends observed in household surveys, in particular, a marked fall in inequality between 2001 and 2014. The monotonic decrease of earnings growth goes only until the 90 percentile above this point the trend is reverted, which is in line with evidence based on Personal Income Tax data. J-Divergence shows that the share of inequality explained by the top 1% and 0.1% rose since 1995 43.1% and 90.1%, respectively. Similarly, the share of inequality explained by those with more than high school increased 29.5% in the same period.

The paper is organized into eight sections as follows. Section 2 discusses the main aspects of the dataset used on this note, in comparison to other distributive studies. Section 3 defines the indicators applied in the analysis. Section 4 provides the details about the data construction process. Section 5 discusses measurement issues on income distribution (such as earnings vs. hourly earnings, missing values, null values and the role of employment on earnings inequality), calculates the standard inequality indexes and plots growth incidence curves and the Lorenz curve. Section 6 applies information theory based decompositions between and within groups. Section 7 disentangle the effects of specific top income and

education groups into inequality changes exploring J-Divergence index properties. The last section concludes.

2. Background of RAIS based Distributive Studies

Most of the analyses on Brazilian income distribution is based on household surveys, in particular the *Pesquisa Nacional de Amostras a Domicílio* (PNAD – IBGE), the main Brazilian National Household Survey. However, RAIS has some advantages. First, it allows combining formal workers and firms' information to understand wage inequality determinants. In particular, the incorporation of individual firms fixed effects explains the bulk of earnings distribution levels and changes (Alvarez et al. 2017). Second, it is the only nationwide data source available with long spells of panel data. This longitudinal aspect allows studying the mobility of workers across sectors and individual firms as well as the life-cycle profile of these characteristics (Machado et al. 2017). Third, RAIS also offers the possibility of analyzing short run employment and wage dynamics because it contains information on a monthly basis – used in Brazil - that allows aggregation to higher time measurement periods - like a year used in most countries⁵. This may facilitate international data comparisons since the measurement unit varies across countries. Fourth, RAIS provides a unique perspective on certain policy related issues. The evaluation of legal employment quotas for People With Disabilities (PWD), and for the youth that requires certain shares of firms employment allocated for these groups is only possible using the establishment as the unit of information and analysis (Neri et al. 2003). RAIS also allows to measure how binding are minimum wages in the bottom of formal employment earnings distribution (Engbom and Moser 2017). Finally, RAIS unlike other data sources does not have top coding which permits to measure wages at the very upper tail of earnings distribution. Nevertheless, RAIS does not include the informal sector, which is very large in Brazil and mostly misses wages at the lower end of the distribution. Employers and top earners that constitute a juridical person for tax purposes are also not in RAIS. With these caveats in mind, we note RAIS has been rarely employed in studies of the Brazilian income inequality until recent years.

Our calculations over PNAD present in Figure 1 show a fall of the Gini of *per capita* income, the most widely used measure, since 1993. However, the bulk of inequality reduction happened between 2001 and 2014. A similar pattern emerges in the concentration index of individual labor income in RAIS (see Figure 2). A second point to notice in the graph is that the fall of *per capita* income from all sources Gini index in the 2001 to 2014 period is more pronounced than the corresponding fall of individual labor concentration index. One possible explanation is that the fall of correlation between schooling of heads and spouse from 0.73 to 0.61 between 2001 and 2015 reinforces the *per capita* income but not that of individual labor earnings. Another possibility is that the expansion of other income sources such as social security benefits and conditional cash transfers is behind this difference (Barros et al., 2006; Kakwani et al., 2010).

3. Inequality Analysis

We briefly describe the inequality measures and decomposition we perform in the paper. Readers familiar with them can skip this section without prejudice. Further details are in the Appendix.

⁵A small exercise for Great Rio in 2015 shows that Gini of monthly earnings are 30% higher than those for annual earnings this include both sources of variability changes of employment and of real wages within the 12-month period.

3.1. Inequality Indexes

Gini Index

The Gini is an inequality index, corresponding to the ratio between the mean absolute deviations of the incomes of all the people in the sample and twice the mean income. N is the population size. Once there are $\frac{N(N-1)}{2}$ distinct pairs of people in the sample, Gini's formula is:

$$Gini = \frac{1}{\mu N(N-1)} \sum_{i>j}^N \sum_j^N |x_i - x_j|$$

where x_i, x_j is individual earnings for two generic and distinct individuals i and j while μ is overall mean income. This formula yields the polar cases:

Perfect Equality: when all individuals have the same income, $x_i = \mu \forall i$, the sum above is equal to zero and Gini is also equal to zero.

Perfect Inequality: when one individual has all the wealth ($N\mu$), we have $N - 1$ pairs with absolute deviations equal to $N\mu$, while the rest of the pairs have null deviations. Therefore, Gini is equal to one.

The fact that the Gini index ranges from 0 to 1 makes its interpretation simpler. The direct calculation of the Gini Index from the Lorenz Curve is another explanation for its popularity. However, since the Gini Index is not decomposable, we complement the analysis by using the Theil Indexes.

Theil Indexes (Theil 1967, Bourguignon 1979, Shorrocks 1980, Foster 1983, Ramos 1993)

The Theil-T index is defined by the following formula:

$$T = \sum_{i=1}^n \frac{x_i}{N\mu} \log \frac{x_i}{\mu}$$

The second Theil measure of inequality is Theil-L index, defined by the following formula:

$$L = \sum_{i=1}^n \frac{1}{N} \log \frac{\mu}{x_i}$$

while in Theil T the inequality factors of weighting within the groups are the share of retained income, in Theil L the inequality factors of weighting within the groups are their respective share of population.

J-Divergence

J-divergence equals to the sum between two Theil inequality measures (T + L):

$$J = \frac{1}{N} \sum_{i=1}^N \frac{x_i - \mu}{\mu} \log \left(\frac{x_i}{\mu} \right)$$

This is another measure based on information theory that relates shares in population with shares in income and evaluates the level of dissonance between both distributions. While the Theil-T departs from population shares and calculates the information dissonance with income shares distribution, the Theil-L runs in the opposite direction from income to population shares. The J-Divergence takes a more neutral position taking the sum of both directions. This measure is known with different names such as *symmetric Kullback-Leibler divergence*, *symmetric relative entropy*, *symmetric Theil measure* or *J-divergence*, in honour of Jeffreys (1946) seminal article. Given this symmetry and other decomposition properties - described next - we choose to express most of our results in terms of the J-Divergence.

The Dual of Theil-T

The dual concept allows comparisons between different inequality measures. Keeping the scale from 0 to 1. And allowing a direct analysis of the introduction of a new proportion of null values in the original income inequality measure. In the case of the Theil-T it can be shown that:

$$T_2 = T_1 - \ln(1 - \phi)$$

where T_1 and T_2 are initial and final values of the Theil-T index before and after adding a ϕ proportion of new null values. Since the Theil-L and the J-Divergence do not admit null values, they also do not admit a Dual measure.

3.2 Within and Between Groups Decompositions

This framework attempts to identify the main structural determinants of inequality. We explore a step further quantifying the close causes of its evolution by performing a standard inequality decomposition exercise among k-groups of a given characteristic such as education, for example:

Theil-T, Theil-L and J-Divergence indexes decompositions⁶

$$T = T_e + \sum_{h=1}^K Y_h T_h$$

where π_h is the proportion of group h in the total population. $T_e = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h}$ is the Theil-T between

groups and $\sum_{h=1}^K Y_h T_h$ is the income weighted average of intra-groups Theils.

The first term of the expression above corresponds to the "between groups" component $T_e = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h}$ while the second term $\sum_{h=1}^K Y_h T_h$ corresponds to the income weighted "within groups" component. We will address these components for subgroups arbitrarily defined according to workers' characteristics (gender, race, age, schooling and region) and firms' characteristics (sector of activity, legal nature of the firm, firm size and firms specific effect). T_e / T is the gross contribution of a certain characteristic to inequality measured by the Theil-T. The Theil-L index can be decomposed in a similar fashion.

$$L = L_e + \sum_{h=1}^k \pi_h L_h$$

Hence, J-Divergence that is the sum of Theil-T and Theil-L can be written as:

$$J = T_e + L_e + \sum_{h=1}^k Y_h T_h + \sum_{h=1}^k \pi_h L_h$$

In the decomposition formulas for the three information theory based inequality indicators presented above, each group has between and within components. Meaning there are differences between income and population shares for each group and also differences within these groups. The standard decomposition

⁶ See Appendix for a step by step derivation of this decomposition.

analysis relies on the sum of all between groups distribution dissonance terms to evaluate their relative contribution to total inequality.

3.3 J-Divergence Specific Groups Decomposition

Besides allowing the usual decomposition between and within groups, the J-Divergence measure yields a non-negative contribution of each individual, or specific groups of individuals in total inequality⁷. Why do we care about specific groups and not only variables? Because, for example, we would like to see how much top 1% incomes, or people with completed higher education contribute to overall inequality measures. Or in the limit we would like to know how much a single person - say the richest person alive - explains overall inequality. This contribution considers each particular group between and within components.

To be sure, departing from the last formula above, instead of summing all groups between components as in the traditional gross contribution analysis, we choose a specific group among k groups and compute its respective particular overall inequality impact picking both between and within respective components. As opposed to other measures derived from information theory such as Theil-T and Theil-L, individual groups contribution to this measure is always greater or equal to zero. This property makes total inequality equals to the simple sum of non-negative individual divergences. This tool allows to go beyond impact of characteristics, and assess the direct impact of specific groups of this characteristic in total inequality.

Overall, we will develop most of the analysis in terms of the J-Divergence measure given its enhanced additive decomposability properties⁸. When we assess the impact of specific groups, such as individuals with higher education or in the top percentile, we take advantage of the J-Divergence additive groups criteria. We will also use J-Divergence in the usual between and within groups' decomposition. In these cases, we also present in the tables the other two Theil indicators to allow visualizing the construction of the J-Divergence measure and to test the robustness of the results found using more widespread measures. We will assess the contribution of different characteristics and groups in 2015 and to the change observed between 1994 and 2015.

4. Data

This research uses RAIS (*Relação Anual de Informações Sociais*), a matched employer-employee dataset provided by the Brazilian Ministry of Labour. It constructs a data set covering the universe of the formal labour market in Brazil through restricted-access administrative records with an average of 33 million observations per year from 1994 to 2015.

In Brazil, firms are required to report all the workers formally employed at some point in the previous calendar year and each worker is identified by a unique number (PIS, *Programa de Integração Social*), which allows us to follow the employees over time and across firms. Firms also have a unique identifier (CNPJ, *Cadastro Nacional de Pessoa Jurídica*). Thus, our dataset allows us to track workers and firms over time. RAIS contains a set of variables on both firms' and employees' characteristics as well as about the characteristics of the employment contract. Precisely, the information in the dataset includes firm-related variables (sector of activity, size, state, etc.), worker-related variables (gender, age, schooling, etc.) and job-related variables (earnings, occupation, weekly hours of work, etc.).

In this note, we restrict the analyses at those employment contracts that were active on December 31. In case of more than one employment, we select that job with the higher salary (in minimum wages). We calculate the real earnings in December 2015 by multiplying the variable "*wage in December (in*

⁷ See also Rohde 2016, Hecksher et al. 2017 and the Appendix.

⁸ Except when we want to incorporate zeros we use the dual concept of the Theil-T. Since it does not exist for the Theil-L and J-Divergence measures. Another advantage of the dual is to keep the domain of the indicator in the 0 to 1 interval.

minimum wages)" and the value of the minimum wage in each year, and using the INPC (*Índice Nacional de Preços ao Consumidor*) as the deflator. This data is available since the beginning of the series.⁹

Wage Measurement

We chose the start year of our data in 1994 because it is the earliest in which we have information about all the variables that will be used in the analysis. Also, it is the period after the stabilization of inflation in Brazil, which would introduce extra measurement error in the analysis. On the other hand, 2015 is the most recent year which we have access to the RAIS data set.

Before 1999, earnings in RAIS were only expressed using Minimum Wages as a numeraire. After that, one may opt between this and nominal earnings expressed in Brazilian Reais. Figure 3 presents inequality measures for 2015 using these two income units possibilities. The two are very similar, which suggests that conclusions are not affected by the concept used.

Missing Values

The individual earnings data present 3.04 percent of reported zeros, which is not allowed according to Brazilian Minimum Wage legislation. We treat these zeros as missing values. The share of zeros falls from 4.83 percent in 1994 to 3.7 percent in 2015 (See Figure 4). Nonetheless, we show that the Theil-T inequality measure that incorporates the zeros (through the dual concept explained in the previous section) remains very similar.

Hourly Earnings

Another possibility is to express inequality in terms of hourly-earnings, instead of total earnings. One may argue that hourly-earning is more relevant than total earning. However, reported hours in RAIS correspond to contractual hours and assume mostly the same value for all observations within the same firm. In Figure 5, we calculate the Theil-T index using both measures. The 2015 inequality level measured with Theil-T rises 28.3 percent with hourly-earnings (0.597/0.466), while the 1994 to 2015 inequality reduction rises almost 10 percentage points, from 17.1 percent (0.099/0.565) to 27.1 percent (0.222/0.819) when we use the latter concept. Between 2001 and 2014, a period of falling inequality, the difference between income concepts amounts to 4 percentage points. Most of the hourly-earnings inequality reduction happens just at the start of the series, but the trends in the two series are almost parallel, as Figure 5 shows.

5. Comparisons between Inequality Measures: Levels and Trends

Mean Growth and Inequality Trends

Before assessing inequality of positive earnings distribution, it is worth addressing mean and dispersion of earnings growth together with formal employment growth (see Figure 6). Between 1994 and 2015, mean earnings grew 29.6 percent in real terms while formal employment grew 107 percent, amounting to a 165.2 percent growth in terms of the total mass of formal wages earned. This means that of the total increase in formal earnings, three quarters are due to formal employment growth. If we subtract the total Brazilian population growth, 34.4 percent according to the PNAD (National Household Survey), the cumulative growth of the earnings mass expressed on a *per capita* basis is 97.4 percent.

An alternative way is to look at the share of formal employees relative to the whole population. In the 1994-2015 period, this share has increased 54 percent, changing from 14.5 percent to 22.3 percent. Figure 7 and Table 1 presents the evolution of standard inequality measures applied to strictly positive earnings according to RAIS in the 1994-2015 period, in which the Gini reduced from 0.547 to 0.472. This trend is also verified for the Theil-L and the Theil-T indexes, and hence the J-divergence. From 2001 onwards, especially until 2014, there is a clearer inequality downward trend and it may be advisable also to consider this period of analysis. For example, when we look at J-Divergence all of the inequality fall

⁹ If we were to use earnings data expressed in Brazilian Reais (R\$) the series would start in 1999.

observed from 1994 to 2001 happened in the first three years. Brazilian inflation fell sharply with the launch of the Real stabilization Plan occurred in mid-1994 but inflation was still falling in the 1994 to 1996 period. This may affect inequality assessment especially when we consider monthly earnings as it is the case in Brazil.

Growth Incidence

Figure 8 plots cumulative growth curves across the 1994 to 2015 period, from the bottom vintile to the top 0.1 percent, yielding growth in the bottom 5 percent of 364.3 percent falling monotonically as we approach the top decile when it reaches 12.27 percent then there is a reversion of this trend reaching growing monotonically as we approach the top 0.1% where growth is 35.31 percent. Zooming in, we separate the growth rates in the lower from the top parts of the distribution. It is clear the compression of inequality up to the top decile and the decompression from this point onwards to the very top of the income distribution (see Figures 9 and 10). The two lowest vintiles in the formal sector are directly affected by the real minimum wage hikes occurred in this period. The value of the minimum wage in 2015 was R\$ 788, situated in between earnings levels in the first two vintiles, R\$ 544 and R\$ 812, respectively.

Lorenz Curves

We start with the most general representation of inequality provided by the Lorenz curve. Figure 11 presents the Lorenz curve in percentiles from 1995 to 2015 in evenly distributed five-year intervals. The curves moved inwards over the years suggesting a continuous earnings inequality reduction. In order to verify the occurrence of Lorenz dominance across these five-year periods, we plot the difference between these curves, as shown in Figure 12. The Lorenz curves for the pair of years 1995 and 2000 and also 2000 and 2005 crossed themselves in the upper percentiles, while the curves for the 2010 and 2015 interval had a slight cross in the bottom percentiles. Only the curves for 2005 and 2010 did not cross, suggesting a more general inequality reduction in this period. To evaluate the whole 20 years period, we compare the extremes 1995 and 2015 in Figure 13. The data shows that the Lorenz curve for 2015 is more equal than the 1995 one in almost all parts of the distribution, except for the very top percentile.

6. Inequality Indexes and between-within decomposition of Theil

Tables 6 to 15 (not presented in this short version of the paper) show the between-within decomposition, for the Theil-T, Theil-L and J-Divergence indexes, considering the following groups of individual characteristics one at a time: gender (female or male workers), schooling (less than high, high school or more than high school), age groups (workers less than 25 years of age, aged 25-35, aged 36-45 or older than 45), race (Indigenous, White, Black, Yellow, Mullato or Ignored) and region (North, Northeast, Southeast, South or Central-West). In general, the results indicate the predominant role played by the "within" component in explaining the total inequality, for the entire historical series of 1994-2015. However, looking at the "between" effect for the educational categories, we observe a relatively higher contribution of this attribute. For instance, in 1994, schooling explained 24.1% ($=0.262/1.086$) of the total inequality measured by the J-Divergence index, while in 2015 this statistic reached 32.8% ($=0.273/0.832$), see Tables 2 and 6.

We also applied the decomposition of the Theil indexes considering firms' characteristics, such as: size (0-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999 and more than 1000 employees), sector of activity (Agriculture, Cattle and Fishing; Manufacturing and Extractive; Construction and Infrastructure; Commerce, Food and Lodging; Transportation, Communications, Financial; Real State, Defense and Public Administration; Education, Health and Social Services; or Other Social Services, Domestic Services, International Organizations), legal nature of firm (Public; Private; Non Profit; Individuals; International) and specific fixed effects of firms.

Similarly to what we found for several individual workers' characteristics above, the between-within decomposition for firms' characteristics shows a predominant power of the "within" component in

determining the total inequality. Nonetheless, when we look at a highly disaggregated level by considering a firm fixed effect (i.e., each firm being a category itself), the results show a remarkable contribution of individual firms. For the 1994 to 2015 period, the contribution of firms' specific factors explained around 65% of total inequality in each year considered. In 2015, the portion of the total inequality measured by the J-Divergence index explained by the between component reached 64.7% ($=0.538/0.83$), see Tables 2 and 14.

Taken together, our findings suggest that, among several workers' characteristics, the differences in schooling between groups were a primary factor in explaining total inequality in the Brazilian formal labor market. However, the explanatory power of firms fixed effects is even more pronounced, playing the major role in determining labor earnings inequality levels in the Brazilian formal labor market.

Changes

When one looks at the changes observed from 1994 to 2015, the explanatory power of individual firms effect to explain the fall of inequality observed is 64.5% (change of between groups $0.5381 - 0.7024 = -0.1643$ over total inequality change -0.2547%), as Table 5 shows. Its last collums are based on the results of Tables 6 to 15. Applying the same type of analysis across time to different characteristics, we have also found: education (-4.3%), gender (2.55%), age (8.8%)¹⁰, macro-region (1.96%), sector of activity (9.92%), nature of the firm (-2.61% from 1995 to 2015)¹¹ and firm size (3.06%). The specific firm-effect explains around three times more the 1994 to 2015's inequality fall than the join gross contribution of all other characteristics considered.

The other striking result is the increasing impact of education on inequality in this period¹². This concentration effect disappears if one considers a more recent period of analysis. From 2001 onwards, there is a clearer inequality downward trend and it may be advisable to also consider this period. Education explained 33.3% of the marked inequality fall observed assuming the role of the second higher explanatory power to explain inequality change (Lam et al. 2015) . Once again, specific firm effects explain 75.9% of inequality fall occurred between 2001 and 2014. Table 2 also presents the contribution of other variables for the 2001-2015 period.

7. The Contribution of Specific Top Incomes and Educational Groups

One key advantage of the J-Divergence is to go beyond the between/within groups dichotomy, allowing to evaluate the role of a specific group in overall inequality. To be sure, by characteristic we mean schooling, and by group we mean those with completed college education, for example. It includes the impact of education premiums paid to those with university degree, their respective share in the population but also the level of inequality within groups¹³. Tables 16 to 24 present the results opened by all groups for all socio-demographic and firms related characteristics explored in the paper.

As we have seen, the main variables that explain formal earnings inequality fall in Brazil during the 1994 to 2015 period are individual firms, schooling, and age. We focus initially here on the group with more than a high school degree. This group explained by itself, in 2015, 48.7% of total inequality while in

¹⁰ Ferreira et al. (2014) emphasizes the reduction of age earnings premium using PNAD surveys.

¹¹ Nature of the firm, that is if a firm is public, private, etc, contributed to a rise of inequality. Courseil et al (2011) show an increase of market concentration on larger firms using RAIS. Alvarez et al. (2017) shows a growing detachment of earnings and productivity distributions in the manufacturing sector.

¹² If we use a finer schooling division with 9 categories, instead of 3 categories, the contribution of education would rise less than 3 percentage points in 2015 but the positive impact of education in the 2001 to 2015 period would be reverted. Measurement error on personal workers characteristics might influence these results.

¹³ If we are interested only in contributions of groups situated in the top part of the income distribution the Theil -T could be used as well. The Theil-T presents always positive contributions to those above the mean (Morley 1999; Neri and Camargo 1999).

1994 it amounted to 37.6% (Figure 14 and Table 16). That is, there was a relative rise of this category relative impact on overall inequality of 29.5% in this period.

Another application of this J-Divergence property explored here is assessing the role played by top income brackets (or individual income of a single person for that matter) in total inequality. According to RAIS (Figures 15 and 16 and Table 25 based on Table 24) between 1994 and 2015: i) the top 10 per cent rose their share in total inequality from 49.91% to 59.97%, a 20.2% rise; ii) the top 5 per cent rose their share in total inequality from 41.4% to 52.2%, a 26.2% rise; iii) the top 1 per cent's share rose from 19.28% to 27.57%, a 43.1% rise. iv) the top 0.1 per cent's share rose from 3.74% to 7.13%, a 91% rise. The concern with top income shares has been increasing around the World (Piketty 2014). The Brazilian case assessed here is curious because it demonstrates that in spite of overall formal earnings inequality fall according to most measures there was an increasing concentration at the very upper part of earnings distribution.

8. Conclusions

The assessment of income inequality normally uses household surveys. More recently, there was a series of papers based on Personal Income Tax (PIT) records and also combining these two types of data sources. However, Brazil also has a long series of establishment level administrative records seldom used in distributive studies. The best example of these microdata sets is RAIS (*Registro Anual de Informações Sociais*) source collected by the Labour Ministry with an average of 30 million observations gathered per year in the last two decades.

This note describes the evolution and the close causes of formal earnings inequality in the Brazilian formal sector from 1994 and 2015 using RAIS. First, we show that earnings distribution changes observed in RAIS reveal a marked inequality fall that is also observed in other more usual measures of inequality extracted from household surveys. For example, the Gini of labor earnings in RAIS fell 12.5% between 1995 and 2015, while the concentration index obtained with PNAD survey fell 19.3% in the same period.

Second, unlike other data sources, RAIS does not have top coding, which permits to measure wages at the very upper tail of earnings distribution. The paper show that in spite of overall inequality fall, the monotonic decrease of earnings increase goes until the 90 percentile and raises specially above the 95 percentile. This concentration increase goes in the same direction as PIT based measures and deserves further scrutiny.

Third, we use standard inequality decompositions applied to the J-Divergence index to understand the main close determinants of inequality. Schooling sticks out among other characteristics explaining 32.8% of total inequality in 2015. The same statistics for individual firms effects reach 64.7%. This means that the gross explanatory power of individual firms to explain inequality in the Brazilian formal labor market is almost twice the one for education. We also explore the change of inequality where firms appear as the main driving variable.

Finally, the paper also explores J-Divergence properties that allows to see beyond the gross contribution of different variables and to capture the relative role played by specific groups. We apply it to isolate the role of top incomes. Our results reveals that since 1995 the share of inequality explained by the top 10%, 1% and 0.1% incomes rose 20.2%, 43.1% and 91%, respectively. Similarly, in spite of falling mean schooling returns, the share of inequality explained by those with high school diploma rises 29.5% in the same period.

References:

- ALVAREZ, J.; BENGURIA, F.; ENGBOM, N.; and MOSER, C. "Firms and the Decline of Earnings Inequality in Brazil," Working Paper, 2017.
- BACHA, E.; TAYLOR, L. Brazilian Income Distribution in the 1960s: 'Facts', Model Results and the Controversy. *Journal of Development Studies*, vol. 14, n. 3, p.271-297, 1978.
- 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.
- BOURGUIGNON, F. Decomposable, income inequality measures, *Econometrica*, 47, pp 901-920, 1979.
- CARD, David; HEINING, Jörg; KLINE, Patrick. Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, v. 128, n. 3, p. 967-1015, 2013.
- COURSEIL, C., MOURA, R. AND RAMOS, L. Determinantes da expansão do emprego formal: O que explica o aumento do tamanho médio dos estabelecimentos?, *Economia Aplicada* Vol. 15, N.1, pp 45-63, 2011.
- ENGBOM, N. and MOSER, C. 2017, Earnings Inequality and the Minimum Wage: Evidence from Brazil, CESIFO Working Paper no 6393, mimeo, March 2017.
- FOSTER, J. E. An axiomatic characterization of the Theil measure of income inequality. *Journal of Economic Theory*, v. 31, n. 1, p. 105-121, 1983.
- FERREIRA, F. H.; FIRPO, S. P.; MESSINA, J. Understanding Recent Dynamics of Earnings Inequality in Brazil. *New Order and Progress: Development and Democracy in Brazil*, v. 187, 2016.
- FISHLOW, A. Brazilian Size Distribution of Income. *American Economic Review*, vol. 52:2, p. 391-402, 1974.
- JEFFREYS, H. F. R. S. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.*, v. 186, n. 1007, p. 453-461, 1946.
- HECKSHER, M.; SILVA, P.L.N.; COURSEIL, C. *Preponderância dos ricos na desigualdade de renda no Brasil (1981-2016): Aplicação da J-divergência a dados domiciliares e tributaries*. Tese de Mestrado ENCE/IBGE, 2017.
- HOFFMANN, R., *Distribuição de renda, medidas de desigualdade e pobreza*, São Paulo: Editora da Universidade de São Paulo (Edusp), 1998.
- KAKWANI, N. ; NERI, M. ; SON, H. Linkages Between Pro-Poor Growth, Social Programs and Labor Market: The Recent Brazilian Experience. *World Development*, v. 38, p. 881-894, 2010.
- LANGONI, C. G. *Distribuição da Renda e Desenvolvimento Econômico do Brasil*. Rio de Janeiro: Editora FGV, 3ª edição, caps. 3 e 5, 2005 (280 p.).
- MACHADO C., NERI, M. and NETO, V The Gender Gap and the Life Cycle in the Brazilian Formal Labour Market, mimeo, presented at ANPEC, 2017.

- 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.
- MORLEY, S. (1999). What happened to the rich and the poor during the post reform period. Santiago, Chile: CEPAL, mimeo.
- LAM, D.; FINN, A.; LEIBBRANDT M. Schooling inequality, returns to schooling and earnings inequality: Evidence from Brazil and South Africa; Wider Working Paper 2015/50, mimeo, 2015.
- NERI, M. *Diversidade*. Rio de Janeiro: Editora FGV, 1ª edição, (204 p.), 2003.
- NERI, M. and CAMARGO, J. M. Distributive Effects of Brazilian Structural Reforms. In: Renato Baumann. (Org.). *Brazil in the 1990s: A Decade in Transition*. New York: Palgrave, 2002, v. , p. 262-313
- PIKETTY, T. *O capital no século XXI*. Rio de Janeiro: Intrínseca, 2014.
- RAMOS, L. A Distribuição de Rendimentos no Brasil 1976/85, Ipea, 135 p, 1993.
- ROHDE, N. J-divergence measurements of economic inequality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, v. 179, n. 3, p. 847-870, 2016.
- SHORROCKS, A., The class of additively decomposable measures, *Econometrica*, 48, pp 613-625, 1980
- 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.
- SONG, J., PRICE, D. J., GUVENEN, F., BLOOM, N., & VON WACHTER, T. Firming up inequality (No. w21199). National Bureau of Economic Research, 2015.
- THEIL, H. *Economics and information theory*. Amsterdam: North-Holland, 1967.

Appendix: Information Theory: Inequality Measures and Decompositions (Theil 1967, Hoffmann 1998)

Theil Inequality Measures

Henri Theil (1967) proposed an inequality measure from the entropy of a distribution.

defined by the following formula:

$$T = \sum_{i=1}^n y_i \log ny_i$$

or, alternatively, by

$$T = \sum_{i=1}^n \frac{x_i}{N\mu} \log \frac{x_i}{\mu}$$

Intra and Inter Groups Decomposition

Suppose I have a population with N samples, divided in K groups:

$N = \sum_{h=1}^K n_h$, which n_h is the n° of people in the h-th group. The proportion of the population correspondent to the h-th group would be:

$\pi_h = \frac{n_h}{N}$. Suppose that x_{hi} is the i-th individual income of the h-th group. Thus, total income share of this individual would be:

$y_{hi} = \frac{x_{hi}}{N\mu}$, note that the denominator is the population total income, with μ as the mean income. So, the share of the total income retained by the h-th group is:

$Y_h = \sum_{i=1}^{n_h} y_{hi}$, that is, adding the share of total income retained by the individuals within group h.

We have Theil-T Index:

$T = \sum_{i=1}^N y_i \log Ny_i = \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log Ny_{hi}$, Firstly, I'm only first the individuals within the group, and then adding the others until complete all the population.

Adding and subtracting:

(*) $\sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} = \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$ (from left to right, I opened Y_h which is out of the log, as defined

above ($Y_h = \sum_{i=1}^{n_h} y_{hi}$). Thereby, the equation turn to:

$T = \sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^k \sum_{i=1}^{n_h} \frac{Y_h}{Y_h} y_{hi} \log Ny_{hi} - \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$, which I added and subtracted (*) and divided and multiplied for Y_h . Continuing:

$$T = \sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log Ny_{hi} - \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$$

$$T = \sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log Ny_{hi} - y_{hi} \log \frac{NY_h}{n_h} \right]$$

$$T = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log \frac{Ny_{hi}}{\frac{NY_h}{n_h}} \right]$$

$$T = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log \frac{n_h y_{hi}}{Y_h} \right]$$

$$T = T_e + \sum_{h=1}^K Y_h T_h$$

Where, $T_e = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h}$ is the Theil-T between groups and $T_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log n_h \frac{y_{hi}}{Y_h}$ is the Theil-T intra groups. Therefore $\sum_{h=1}^K Y_h T_h$ is the weighted average of intra-groups Theils.

T_e / T is the Contribution of a certain characteristic to inequality measured by the Theil-T.

Similarly, we can show that the Theil-L can be decomposed as between groups (L_e) and within groups components:

$$L = L_e + \sum_{h=1}^k \pi_h L_h$$

$$\text{where } L_e = \sum_{h=1}^k \pi_h \log(\pi_h / Y_h) \text{ and } L_h = \frac{1}{n_h} \sum_{i=1}^k \pi_h \log(Y_h / (n_h y_{hi}))$$

Hence, J-Divergence can also be expressed in terms of its within and between groups components. By its turn each of these components can be expressed in terms of the sum of Theil-T and Theil-L respective components:

$$J = T_e + L_e + \sum_{h=1}^k Y_h T_h + \sum_{h=1}^k \pi_h L_h$$

J-Divergence Group Decomposition (Jeffreys 1946; Rohde 2016; Hecksher et al 2017):

The J-Divergence measure allows to gauge the contribution of specific groups of individuals in total inequality. How is it done? In the within and between decomposition formulas for the three information theory based inequality indicators above, instead of summing all groups between groups component, we instead choose a specific group among k groups and compute its respective contribution from both between and within components.

At this point lies a comparative advantage of the J-Divergence. As opposed to other measures derived from information theory such as Theil-T and Theil-L, individual's contribution to this measure is always greater or equal to zero. In Figure 17, we see that while the Theil-T receives negative contributions from individuals below the mean and the Theil-L receives non-negative contributions for those above the mean while in the J-Divergence, these individuals contributions are always non negative. This property makes the simple sum of individual divergences equals to total inequality, allowing analyzing the direct impact of specific groups' in inequality.

The Contribution of a Characteristic and Group to Inequality Level (and Growth)

The contribution of a given characteristic to inequality level and change exemplified initially by the J-Divergence of a given characteristic is:

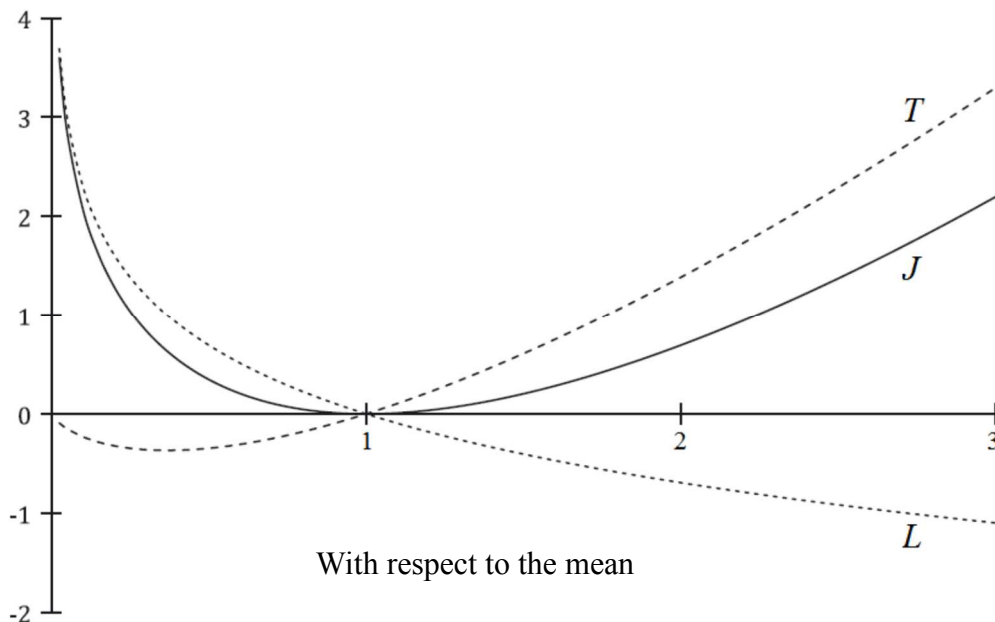
$$\text{Gross Contribution } J = J_{et} / J_t$$

$$\text{Share of Gross Contribution Change in total Change} = \Delta(J_{et}) / \Delta(J_t)$$

where $J_e = T_e + L_e$ and $J = T + L$

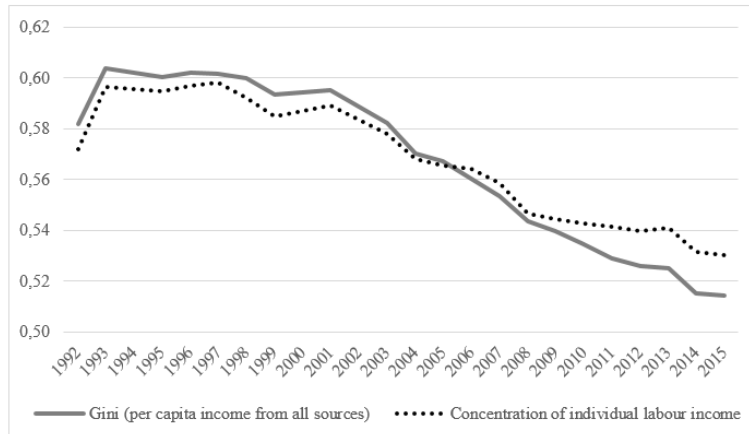
The two equations above says that the relative contribution of a given characteristic – say schooling - to inequality level in a single point in time (change across time) is given by its between component level (change) divided by initial total inequality. Since in the J-Divergence the same additive decomposability is applied do specific groups – say individuals with higher education diploma - exactly the same idea can also be applied to assess the gross impact of a specific group of a given characteristic to inequality.

Contributions to Inequality according to Income: Theil-T, Theil-L & J-Divergence



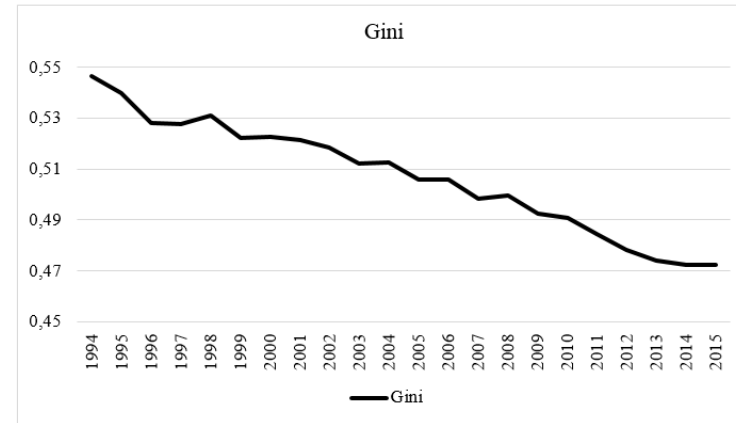
List of Figures

Figure 1: Inequality (Gini Index) in Household Surveys 1994 – 2015



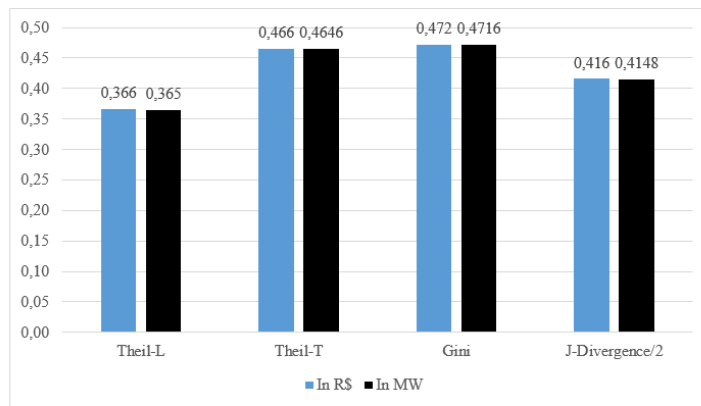
Source: Author's calculation over PNAD microdata.

Figure 2: Evolution of the Gini Index in RAIS 1994 - 2015



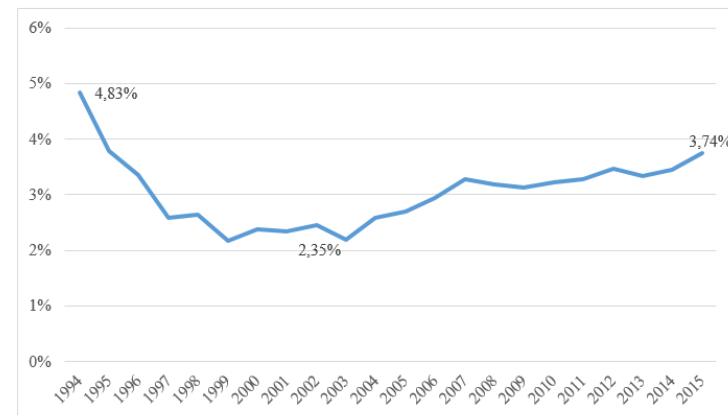
Source: Author's calculation over RAIS microdata.

Figure 3: Earnings Inequality During 2015 in R\$ and in Minimum Wages (MW)



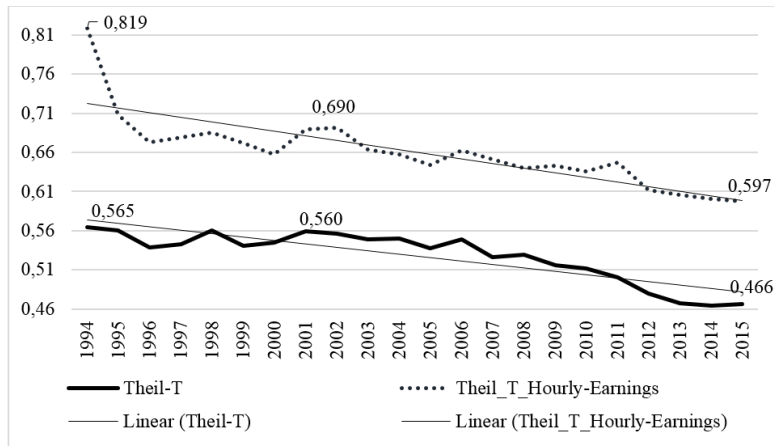
Source: Author's calculation over RAIS microdata.

Figure 4: Share of Missing Incomes (% 0s - Measurement Error)



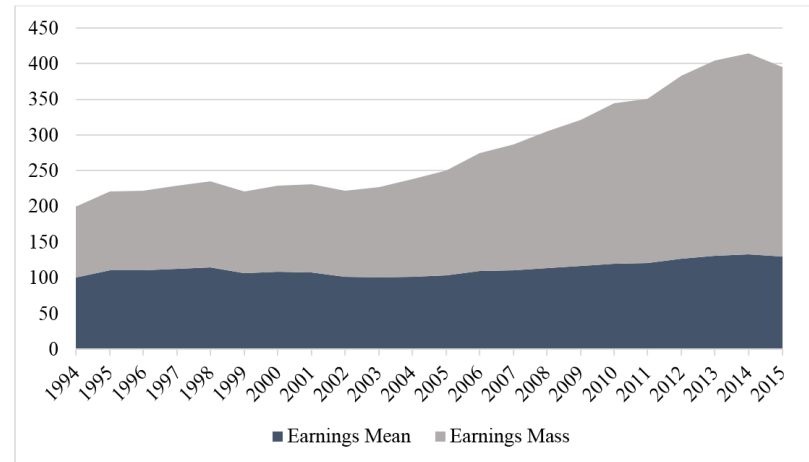
Source: Author's calculation over RAIS microdata.

Figure 5: Earnings versus Hourly Earnings Inequality – Theil-T 1994-2015



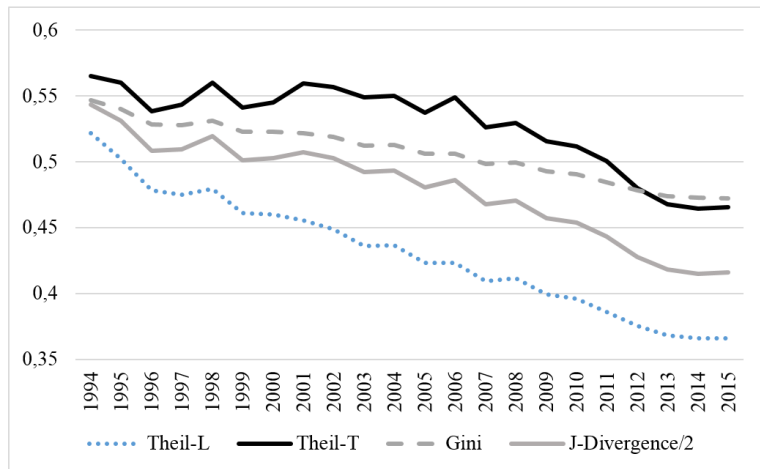
Source: Authors' calculation over RAIS microdata.

Figure 6: Earnings Mean and Earnings – 1994-2015



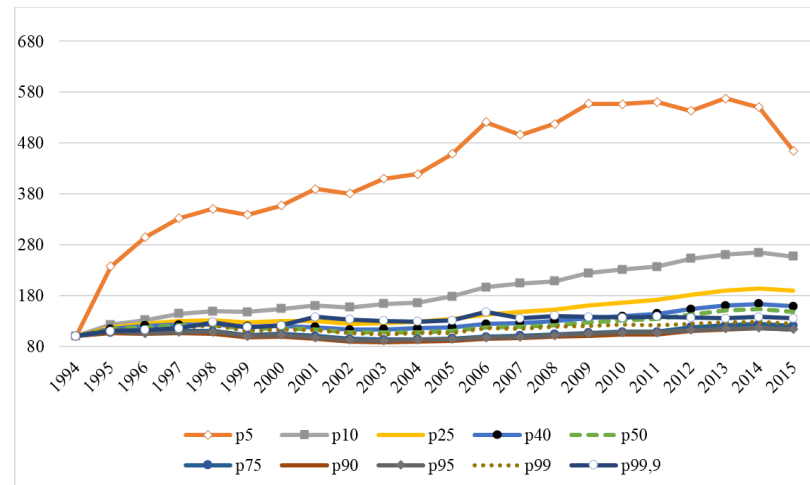
Source: Authors' calculation over RAIS microdata.

Figure 7: Various Inequality Measures Trends 1994 – 2015



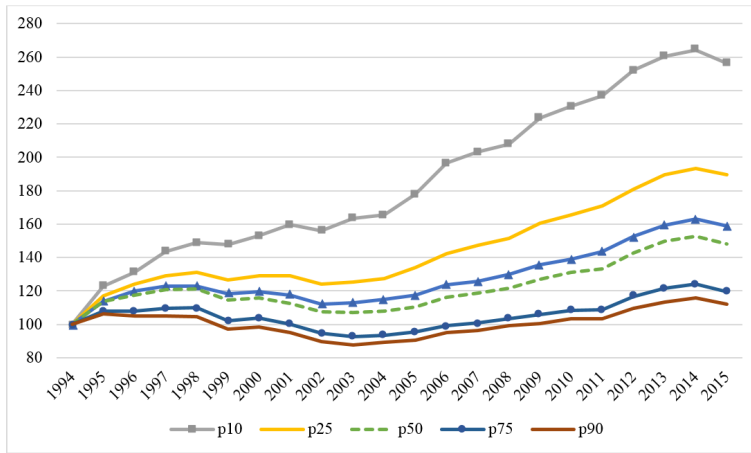
Source: Authors' calculation over RAIS microdata.

Figure 8: Cumulative Growth Curve Across Percentiles -1994 - 2015



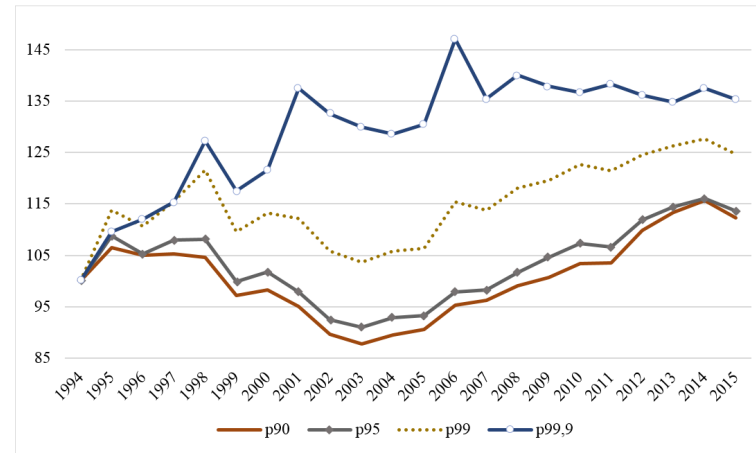
Source: Author's calculation over RAIS microdata.

Figure 9: Cumulative Growth Curve Across Lower Percentiles -1994 - 2015



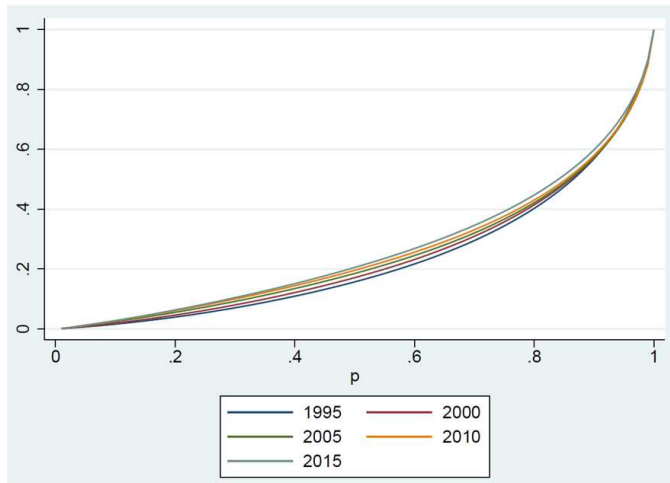
Source: Author's calculation over RAIS microdata.

Figure 10: Cumulative Growth Curve Across Top Percentiles -1994 - 2015



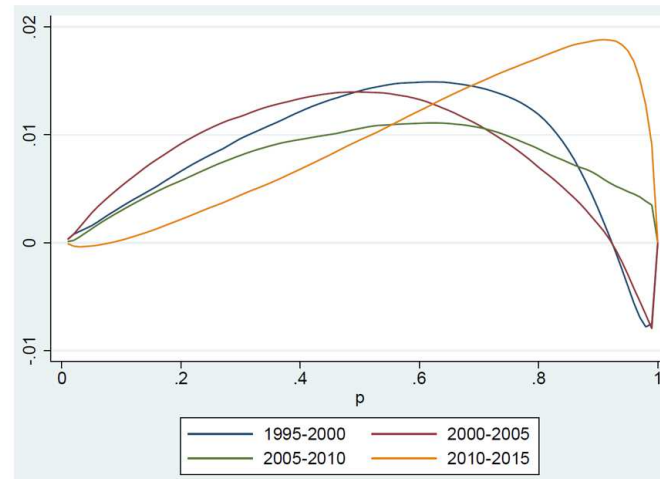
Source: Author's calculation over RAIS microdata.

Figure 11: Lorenz Curves in Five-Year Intervals



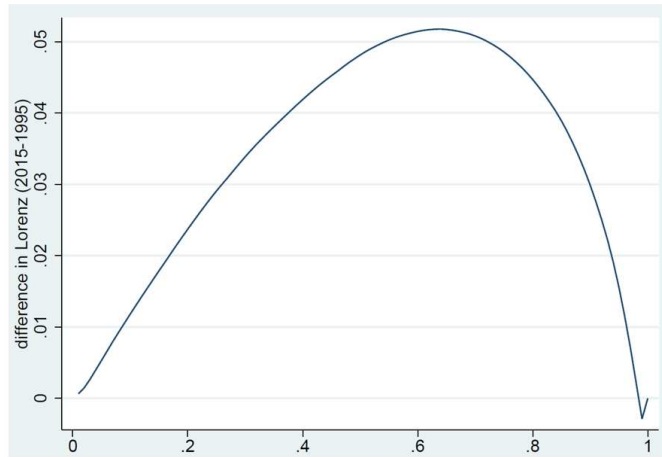
Source: Authors' calculation over RAIS microdata.

Figure 12: Lorenz Curves Differences in Five-Year Intervals



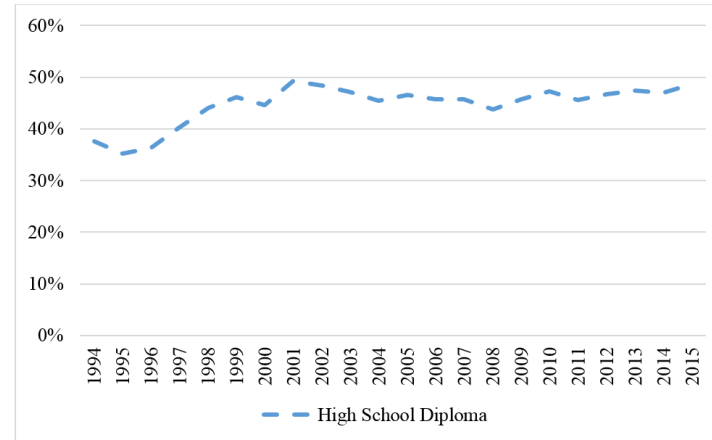
Source: Author's calculation over RAIS microdata.

Figure 13: Lorenz Curves Differences Between 1995 and 2015



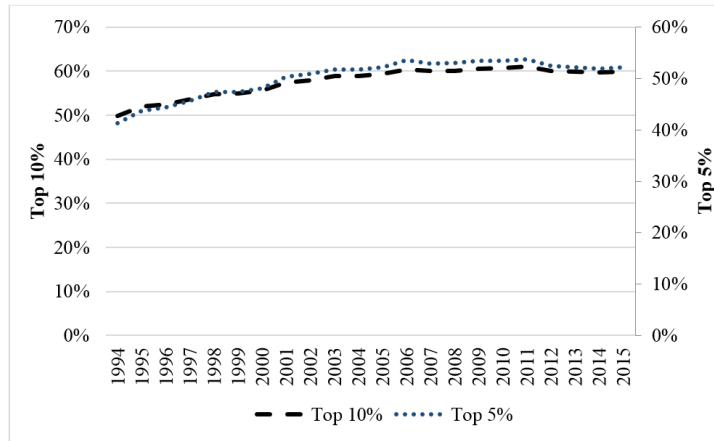
Source: Author's calculation over RAIS microdata.

**Figure 14: Specific Groups Contributions to Inequality: *J-Divergence*
High School Diploma Holders**



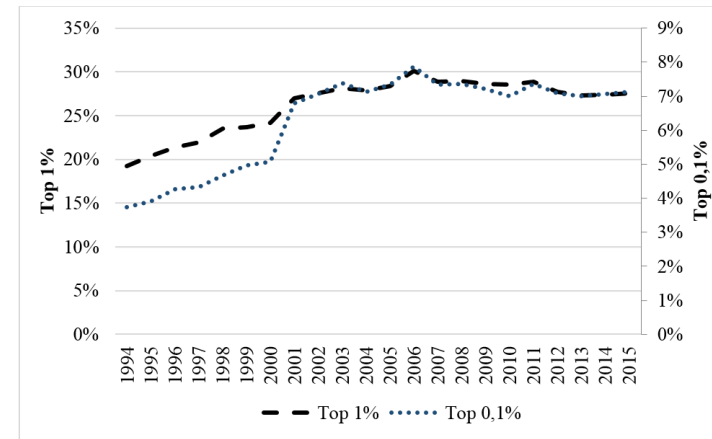
Source: Author's calculation over RAIS microdata.

**Figure 15: Specific Groups Contributions to Inequality: *J-Divergence*
Top 10% Incomes and Top 5% Incomes**



Source: Author's calculation over RAIS microdata.

**Figure 16: Specific Groups Contributions to Inequality: *J-Divergence*
Top 1% Incomes and Top 0.1% Incomes**



Source: Author's calculation over RAIS microdata.

Table 1: Evolution of Inequality Measures of Strictly Positive Values

Year	Gini	Theil-L	Theil-T	J-Divergence	Dual_Theil-T
	-----Stricly Positive Values-----				
1994	0.547	0.522	0.565	1.086	0.432
1995	0.54	0.502	0.56	1.062	0.429
1996	0.528	0.478	0.538	1.016	0.416
1997	0.528	0.475	0.543	1.018	0.419
1998	0.531	0.479	0.56	1.039	0.429
1999	0.522	0.461	0.541	1.002	0.418
2000	0.522	0.46	0.545	1.005	0.42
2001	0.521	0.455	0.56	1.015	0.429
2002	0.518	0.449	0.557	1.006	0.427
2003	0.512	0.436	0.549	0.984	0.422
2004	0.513	0.437	0.55	0.986	0.423
2005	0.506	0.423	0.537	0.961	0.416
2006	0.506	0.423	0.549	0.972	0.422
2007	0.498	0.409	0.526	0.935	0.409
2008	0.5	0.412	0.529	0.941	0.411
2009	0.492	0.399	0.515	0.915	0.403
2010	0.491	0.396	0.512	0.908	0.401
2011	0.485	0.386	0.501	0.887	0.394
2012	0.478	0.375	0.48	0.855	0.381
2013	0.474	0.368	0.468	0.836	0.374
2014	0.472	0.366	0.464	0.83	0.371
2015	0.472	0.366	0.466	0.832	0.372
Mean	0.507	0.431	0.528	0.959	0.41

Source: Authors' calculation over RAIS microdata.

Table 2: J-Divergence Index: Different Characteristics Contributions

Characteristics	Level		Share of Change	
	2015	1994	1994 - 2015	2001-2015
Schooling	32,81%	24,11%	-4,32%	33,33%
Gender	0,96%	1,33%	2,55%	0,82%
Age	10,82%	10,35%	8,80%	10,82%
Race*	11,11%	5,09%	-	-27,67%
Region	1,77%	1,81%	1,96%	7,54%
Sector	8,63%	8,93%	9,92%	8,37%
Firm Legal Status**	8,15%	3,78%	-2,61%	-2,61%
Firm Size	13,62%	11,15%	3,06%	7,65%
Firm Specific Effect	64,70%	64,66%	64,53%	75,86%

* from 2003 to 2015; ** from 1995 to 2015 Source: Authors' calculation over RAIS microdata.