*Earnings inequality in the Brazilian formal sector:*

The role of firms, education and top incomes (J-Divergence)

Marcelo Neri (FGV Social)

Cecília Machado (FGV EPGE)

Valdemar Neto (FGV EPGE)


*** paper: https://www.wider.unu.edu/publication/earnings-inequality-brazilian-formal-sector

Key Reference: Alvarez, Benguria, Engbom and Moser (2018)

RAIS Data

- **RAIS** (Relação Anual de Informações Sociais), a matched employer-employee data set provided by the Brazilian Ministry of Labour. It constructs a longitudinal 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 today.

- Allows us to match workers and firms and to track them over time.

**Background of RAIS based Distributive Studies** - Most of the analyses on Brazilian income distribution is based on household surveys in particular Pesquisa Nacional de Amostras a Domicilio (PNAD – IBGE), the main Brazilian National Household Survey. However, RAIS has a few advantages. First, it allows combining workers and firms information to understand formal wage inequality determinants. In particular, the incorporation of individual firms fixed effects explains the bulk of earnings distribution levels and changes (Alvarez et all 2017; Machado et all 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 that allows aggregation to higher time measurement periods. 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 unit of analysis (Neri et al. 2003). RAIS also allows to measure how bidding are minimum wages in the bottom of formal employment earnings distribution (Engbom and Moser 2017). On the other extreme, RAIS unlike other data sources does not have top coding which permits to measure wages at the very upper tail of earnings distribution. And last, and perhaps most importantly, it allows to check the robustness of other types of data sources mentioned. In spite of all these advantages, RAIS was very little used up to know on understanding levels and changes in Brazilian income distribution.
Lorenz Curves Levels in Five-Year Intervals

Source: RAIS microdata

Lorenz Curves Differences in Five-Year Intervals

Source: RAIS microdata
Lorenz Curves Differences Between 1995 and 2015

Source: RAIS microdata

Formal Labour Market in Brazil: Cumulative Growth Incidence Curve 1994 – 2015

Source: RAIS microdata
Formal Labour Market in Brazil:
Cumulative Growth Incidence Curve 1994 – 2015

• Lower percentiles

The earnings growth is bigger for the poorest of the distribution

Source: RAIS microdata

Formal Labour Market in Brazil:
Cumulative Growth Incidence Curve 1994 – 2015

• Top percentiles

From p90 onwards there is increasing growth

Source: RAIS microdata
**J-divergence measurements of economic inequality**

T & L Theil (1967) indexes can be defined as:

\[
T = \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\mu} \ln \left( \frac{x_i}{\mu} \right) \quad \text{and} \quad L = -\frac{1}{N} \sum_{i=1}^{N} \ln \left( \frac{x_i}{\mu} \right),
\]

Where \(x_i\) is individual income, \(N\) is population size and \(\mu\) is mean income.

**J-divergence** (Jeffreys, 1946) measure is the simple sum of Theil T and Theil L indexes \((J = T+L)\) expressed as:

\[
J = \frac{1}{N\mu} \sum_{i=1}^{N} (x_i - \mu) \ln \left( \frac{x_i}{\mu} \right).
\]

Rohde (2016); Hecksher et al. (2017); Neri and Hecksher (2018)

---

**FORMAL EARNINGS INEQUALITY**

<table>
<thead>
<tr>
<th>Year</th>
<th>Theil-L</th>
<th>Theil-T</th>
<th>Gini</th>
<th>J-Divergence/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Gross Contribution to Inequality J-Divergence**

<table>
<thead>
<tr>
<th></th>
<th>Education*</th>
<th>Gender</th>
<th>Age</th>
<th>Firm size</th>
<th>Sector</th>
<th>Type of Firm</th>
<th>Individual Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levell 2015</td>
<td>32,81%</td>
<td>0,96%</td>
<td>10,82%</td>
<td>13,62%</td>
<td>8,63%</td>
<td>8,15%</td>
<td>64,7%</td>
</tr>
<tr>
<td>Change 2001 to 2015</td>
<td>33,33%</td>
<td>0,82%</td>
<td>10,82%</td>
<td>7,65%</td>
<td>13,47%</td>
<td>-2,61%</td>
<td>75,86%</td>
</tr>
</tbody>
</table>

Source: RAIS microdata
**Key point: J-divergence** implies in shares (Always non negative) for each income-bracket and individuals. It allows to move from variables to specific groups (higher education, top 1% incomes, etc).

Rohde (2016); Hecksher, Courseil and Silva (2017), Neri and Hecksher (2018); Morley (1999)

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).

<table>
<thead>
<tr>
<th>Gross</th>
<th>Contribution to Inequality</th>
<th>J-Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education: Completed College</td>
<td>Firm Size: &gt;1000 Employees</td>
<td>Income Brackets: Top 0.1%</td>
</tr>
<tr>
<td>Level 2015</td>
<td>48.7%</td>
<td>34.8%</td>
</tr>
</tbody>
</table>

Source: RAIS microdata
Specific Groups Contributions to Inequality: J-Divergence: Top 10% Incomes and Top 5% Incomes

- Inequality explained by the top 10%, incomes rose 20.2%

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

- Inequality explained by the top 1% and 0.1% incomes rose 43.1% and 91%
Conclusions (RAIS)

Describes the evolution and the close causes of formal earnings inequality in the Brazilian formal sector using RAIS (33 million observations yearly).

Gini of labour earnings in RAIS fell 12.5% between 1995 and 2015, concentration index from PNAD survey labor fell 19.3 per cent.

RAIS does not have top coding, measure wages at the very upper tail of earnings distribution. In spite of overall inequality fall, the monotonic decrease of earnings increase goes until the 90 percentile, in the same direction not timing as PIT-based measures.

J-Divergence allows to capture the role played by specific groups. We apply it to isolate the role of top incomes. Between 1995 and 2015 the share of inequality explained by the top 10%, 1% and 0.1% incomes rose 20.2 per cent, 43.1 per cent and 91 per cent, respectively. Similarly, in spite of falling mean schooling returns, the share of inequality explained by those with high school diploma rises 29.5 %.

Main close determinants of inequality level. Schooling explains 32.8% of total inequality in 2015. Individual firm-effects reach 64.7%

Main determinants of inequality change. Schooling 33.3% of inequality between 2001-15. Individual firm-effects reach 75.9%

J-divergence measurements of per capita income inequality

Rohde, N. (2016) – Summary: The paper uses a symmetric entropy statistic to study income inequality. The index quantifies the information content of a two-way message that transforms the empirical income distribution into an egalitarian reference distribution, and then back to the original. This allows the measure to be interpreted as an average of n income-to-mean divergences such that the inequality estimate can be broken down into contributions across population subgroups. Various properties of the index are analysed and an application comparing the USA, Germany and Britain is provided. We focus on the sensitivity of inequality to the tails of the income distribution and show that the extreme right-hand tail accounts for a large and generally increasing proportion of total inequality. This result holds even if incomes are measured at the household level, averaged over a 5-year period and taken after government taxes and transfers.

Dominance of the richest in Brazilian income inequality: application of J-divergence to household and tax data Hecksher et al. (2017) ** Summary: The share of the income inequality explained by the 10% richest in Brazil is higher than 50%. Higher in Brazil than for the US (45%), Germany (44%) and Great Britain (41%). Inequality was measured using an index which is still not much used in the socioeconomic literature, the J-divergence. It can be defined as the sum of Theil’s T and L indices, but unlike these and the Gini index, can be easily decomposed as the sum of the individual contributions to the total inequality. Equivalised per capita household total monthly income PNAD were used to estimate the J-divergence from 1981 to 2015, and the corresponding shares of the inequality explained by each vintile of the income distribution. By integrating PNAD and income tax data for 2014, more than 50% of the resulting inequality of adult personal income is driven by the top percentile.
**J-Divergence of equivalent per capita Household (HH) income groups**,

Equivalence Scales:

F income sources (labor, rents, social security, Bolsa Familia etc)

\[ \frac{1}{N} \sum_{i=1}^{N} \sum_{f=1}^{F} Y_{if} \]

Per capita HH Income usual:

\[ \frac{1}{N} \sum_{i=1}^{N} \sum_{f=1}^{F} Y_{if} \]

+ general case:

\[ 0 \leq \theta \leq 1; \theta \text{ is an economy of scale parameter} \]

If \( \theta = 1 \) per capita HH income

If \( \theta = 0 \) Total HH income

If \( \theta = 1/2 \) Equivalized per capita HH income – Square root rule

The 10% richest get % share of J-Divergence:

- 45% in the US
- 44% in Germany
- 41% in the U.K.

Turning Point of Inequality of Household Per Capita Income (marginal income increases that lead to inequality increases)

Not very sensitive wrt lower end of distribution
Are firms effects driving inequality in Brazil?

- formal and top earnings in matched employer-employee records

by Marcelo Neri, Cecília Machado and Valdemar Pinho Neto

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 release of the Demographic Census microdata. Recently, a series of papers have documented inequality based on Personal Income Tax (PIT) records. However, establishment-level administrative records are also available in Brazil, but 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. RAIS depicts formal employment dynamics and wage differentials and is a powerful tool that may complement the evidence presented by other data sources.

This work documents 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 that present somewhat similar trends. For example, the Gini of labour 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.

Breaking-down inequality - Standard inequality decompositions based information theory help us 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.). 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% of the total inequality measured by the J-Divergence index, while in 2015 this statistic reached 32.8%.

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 which consumes quite a lot of degrees of freedom), 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%.

 MAIN FINDINGS:

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.

In 2015, schooling explained 32.8% of the overall formal earnings inequality.

The contribution of firms' specific effects are even more important, explaining more than 64% of the total inequality for each year in the data.

Firms also seems to drive the overall inequality in developed countries such as the U.S and Germany.

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 labour market. However, the explanatory power of firms fixed effects is even more pronounced, playing the major role in determining labour earnings inequality levels in the Brazilian formal labour market.

Inequality 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%. Applying the same type of analysis across time to different characteristics, we have also found: education (-4.3%), gender (2.55%), age (8.8%), macro-region (1.96%), sector of activity (9.92%), nature of the firm (-2.61% from 1995 to 2015) 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 period. This concentration effect disappears if one 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. Once again, specific firm effects explain 75.9% of inequality fall occurred between 2001 and 2015. This means that the gross explanatory power of individual firms to explain inequality in the Brazilian formal labour market is almost twice the one for education. In sum, in the context of inequality change, firms also appear as the main driving variable.

Top Incomes - Besides applying between and within groups decomposition for Theil T and Theil L indexes, we use J-Divergence measures to disentangle the role played by specific categories of different variables. Unlike other data sources, RAIS does not have top coding, which permits to measure wages at the very upper tail of earnings distribution.

IMPLICATIONS ON TOP INCOMES:

In spite of overall inequality fall, 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 (see figures).

J-Divergence analysis of the role played by specific categories show that the share of inequality explained by the top 10%, 1% and 0,1% rose since 1995: 20.2%, 43.1% and 90.1%, 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.