*01.082 Horizontal Inequality, Labor Markets and Education

Labor decompositions, Mincerian and Markovian equations, D in D Measurement Error, Selectivity and Ommited Variable Biases

Marcelo Neri (FGV Social)

*** 7.4 text Drivers of Income Distribution Changes** https://www.cps.fgv.br/cps/bd/curso/Drivers_IncomeDistribution_Neri_Brazill_Updated_GMD.pdf

*****Returns to education and intergenerational mobility**
https://www.wider.unu.edu/publication/returns-education-intergenerational-mobility-and-inequality-trends-brazil-0

Labor Deconstruction

Per Capita Labor Income in the total population can be expressed as:

Total Labor Earnings = Mean Earnings of those with Earnings * Occupied Population

We can continue decomposing each piece of the identity in elements, what helps to understand the relative weight of each labor ingredient.

Mean Labor Earnings = MEAN EARNINGS / HOURs x WORKING HOURS x OCUP/ EAP x EAP/ POP

Average Hourly Wages — Weekly Effort — Occupation in the Economically Active Population (EAP) — Participation Rate

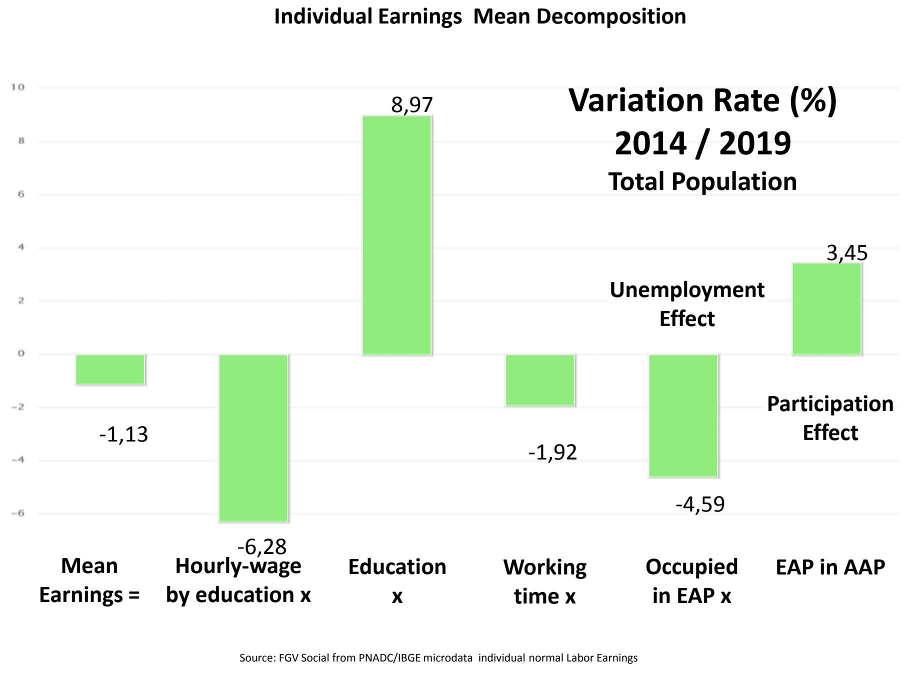
Definitions and Formulas

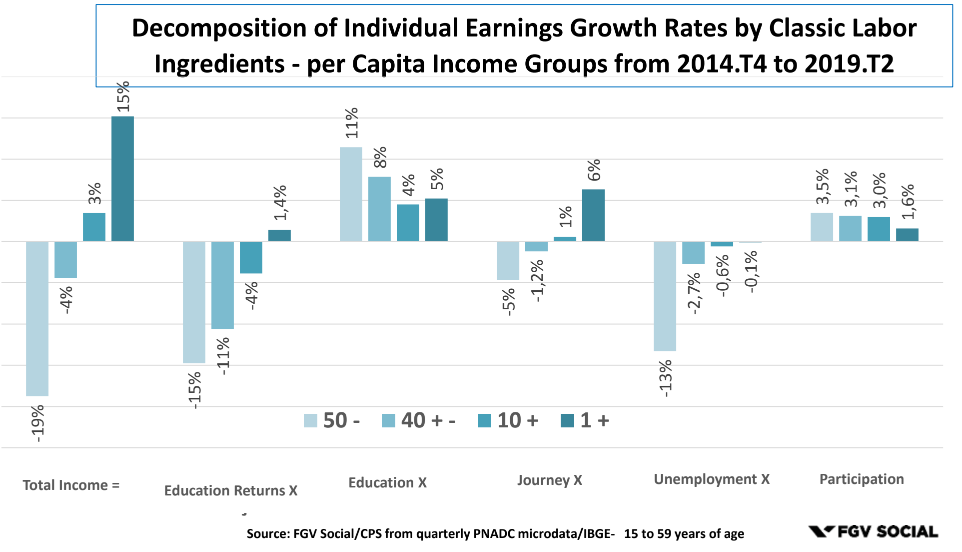
Labor Economics

Occupied population (E): People working
Unemployed population (U): People looking for job but not occupied
Inactive population (I): People not occupied

Active Age Population AAP (*PIA*):
occupied + unemployed + inactive = (E + U + I)
Economically Active Population EAP (*PEA*)
occupied + unemployed (E + U)
Participation Rate: *(PEA) / (PIA) = (E + U) / (E + U + I)*
Unemployment Rate: *(Unemployed) / (PEA) = (U) / (E + U)*
Occupation Rate in *PEA*: *(Occupied) / (PEA) = (E) / (E + U)*

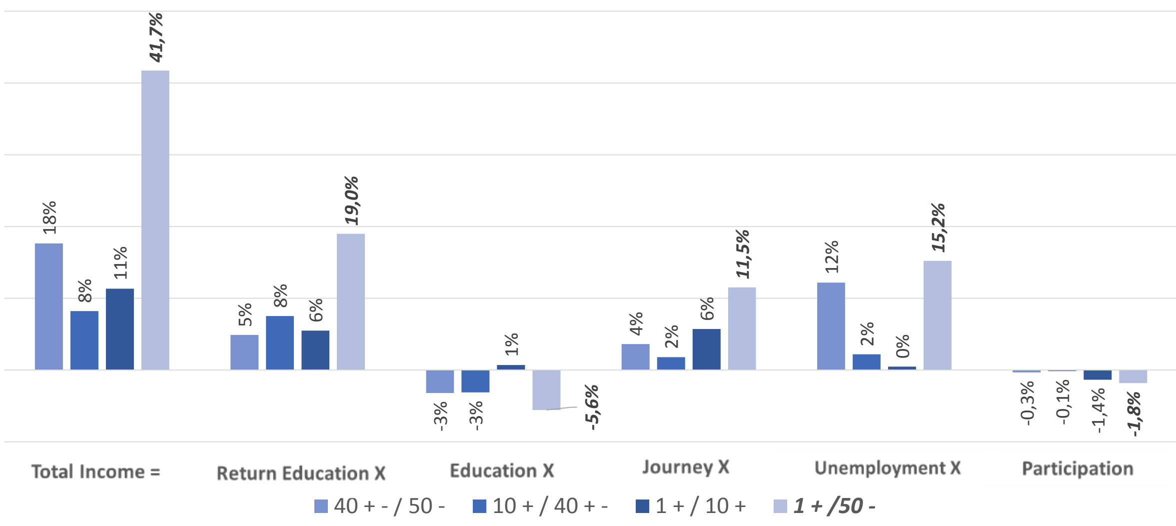
2





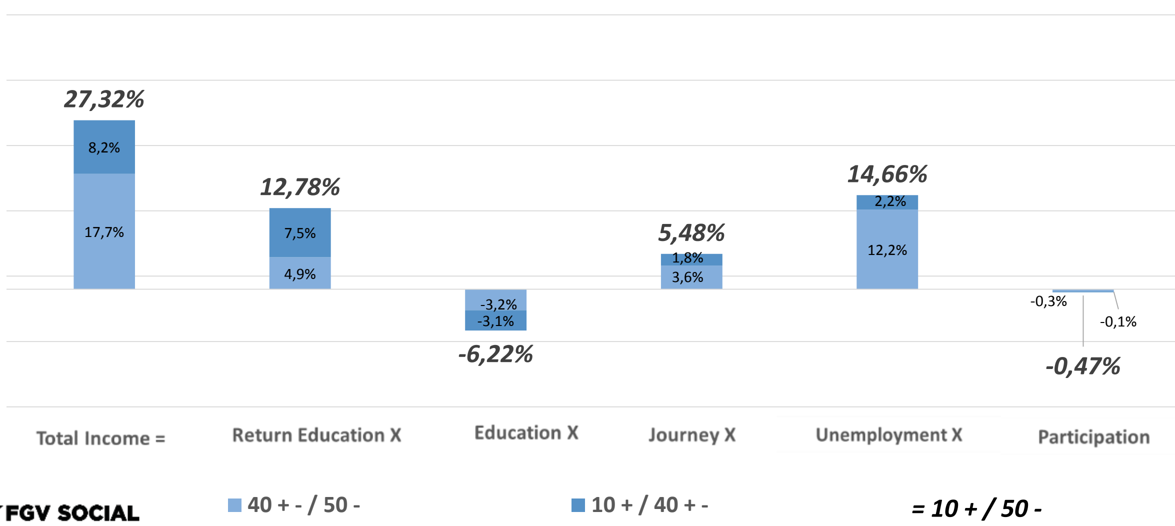
The Escalation of Inequality – What was the impact of the crisis on income distribution and poverty? https://cps.fgv.br/en/inequality

Change in Inequality *1% + / 50%* - and its opened Labor Ingredients for intermediate bands reasons - 2014.T4 to 2019.T2 – 15 to 59 years

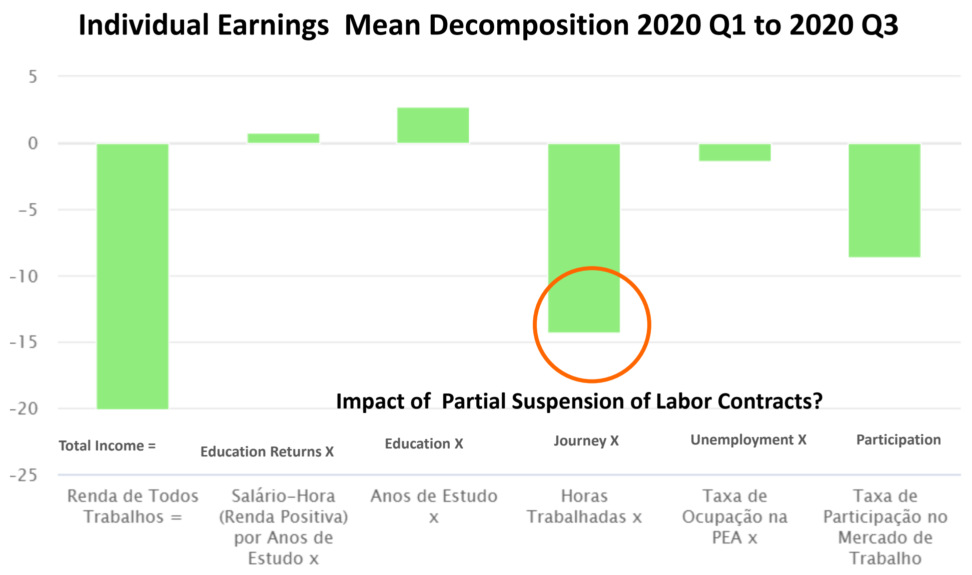


Source: FGV Social/CPS from quarterly PNADC microdata/IBGE. OBS: individual income by per Capita income bands all from work

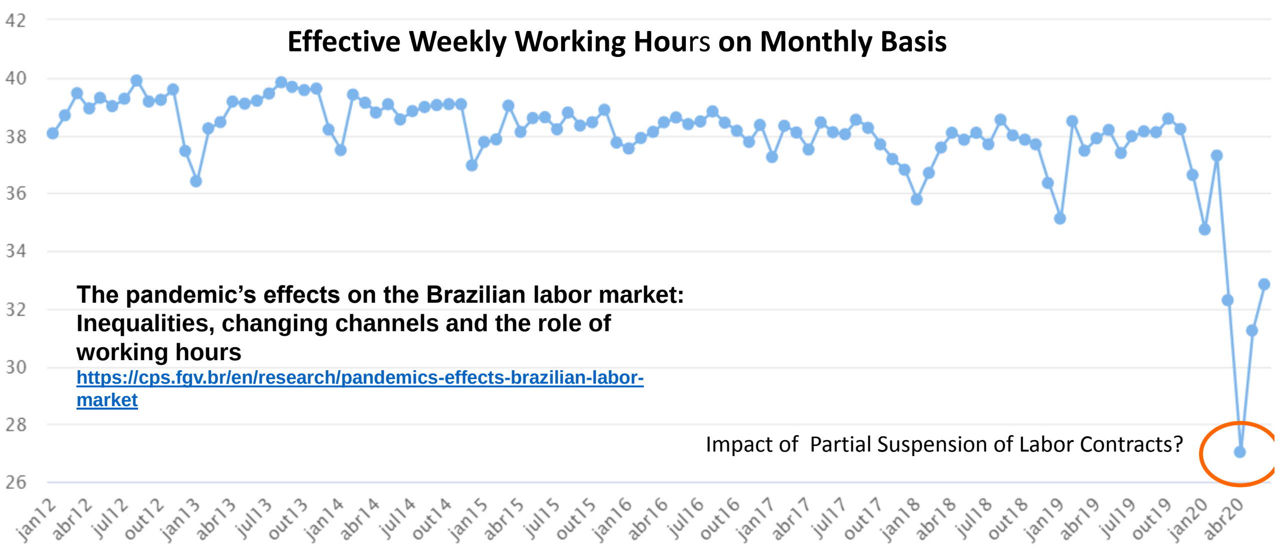
Change in Inequality *10% + / 50%* - and its opened Labor Ingredients for reasons between stacked intermediate bands - 2014.T4 até 2019.T2 – 15 a 59 anos



Source: FGV Social/CPS from quarterly PNADC microdata/IBGE. OBS: individual income by per Capita income bands all from work



Source: FGV Social from PNADC/IBGE microdata individual Effective l Labor Earnings

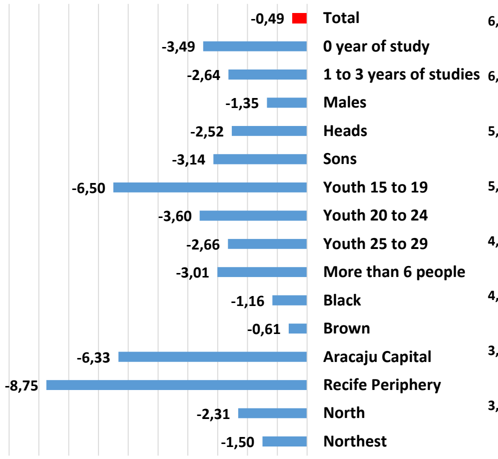


Source: FGV Social from PNADC/IBGE microdata individual Effective Working Hours

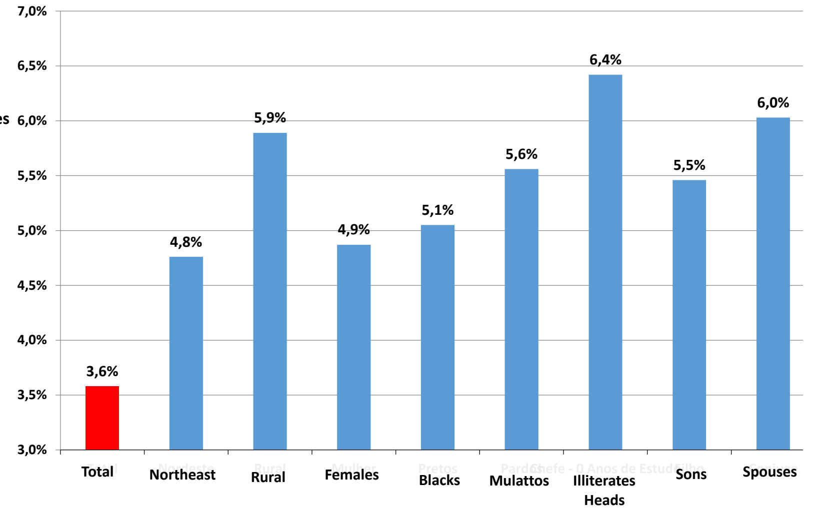
Horizontal Inequality – Individual Income Growth Rate Excluded Groups

Individual Income from Labor - 15 to 59 years (%.)

Annual Variation from 2014 to 2019



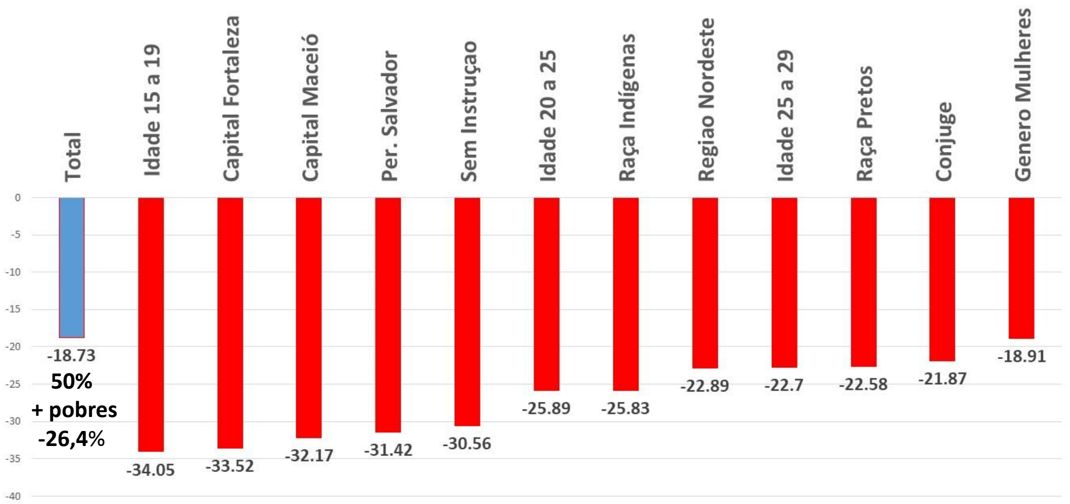
Individual All Sources Incomes Annual Growth Rates - 2001 to 2014



Source FGV Social from PNAD e PNADC/IBGE microdata

***Individual All Labor Effective Earnings Annual Growth Rates* - 2020 Q1 to 2020 Q3**

Quem Perdeu Mais na Pandemia do Covid-19? – Desigualdade Horizonal
Variaçao de Renda Individual Efetiva do Trabalho Real 2020 T1 a 2020 T3



Source FGV Social from PNAD e PNADC/IBGE microdata

Minncerian Model: (Mincer 1974; Lemieux 2006, Card 2001) *01.20

$$y_i = \ln{(Y_i)} = \alpha + \beta S_i + x_i'\gamma + \varepsilon_i$$

where Y_i is the labour income of individual i (we change this metric below) , S_i is the level of education of individual i measured by years of schooling, x_i is a vector of controls and ε_i is an error term.

The Coefficient and Attribute Premium

This is a regression model in the log-level format, that is, the dependent variable, the wage is in logarithmic format and the most relevant independent variable, schooling, is in level format. Therefore, the coefficient β1 measures how much one year more of schooling causes in proportional variation in the wage of the individual. For example, if β1 is estimated at 0.18, this means that each additional year of study is related on average with a wage increase of 18%. This corresponds to the premium of the attribute (or rate of return if the costs were zero). Mathematically, we have:

Deriving, we find that: $(\partial \ln y / \partial \text{ educ }) = \beta_1$

On the other hand, by the chain rule, we have:

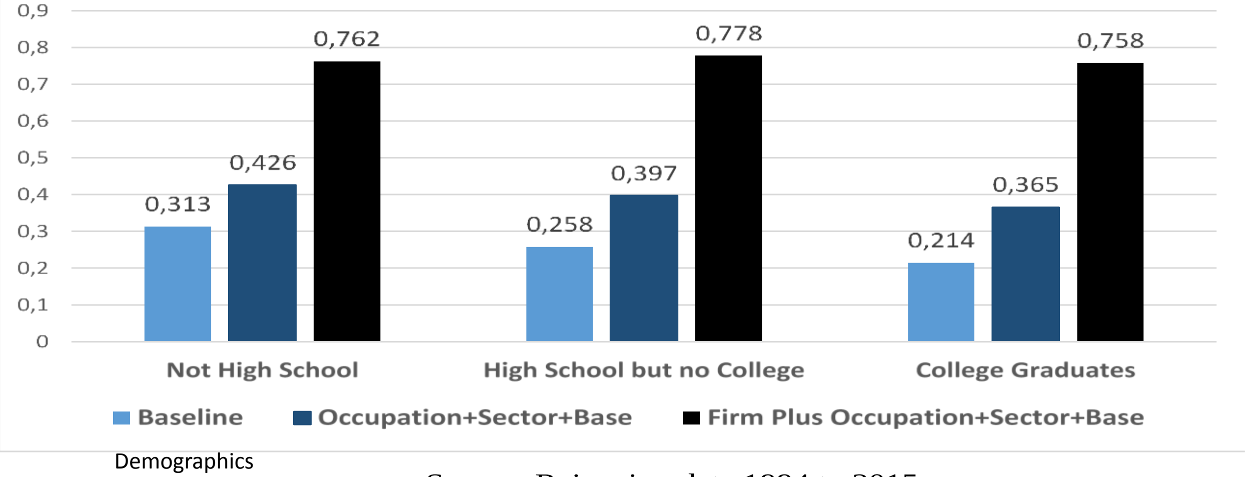
$(\partial \ln y / \partial \text{ educ }) = (\partial y / \partial \text{ educ }) (1 / w) = (\partial y / \partial \text{ educ }) / y)$

Thus, $\beta_1=(\partial y / \partial \text{educ})/ y$, corresponds to the percentage variation of the wage from a increase of one year of study..

The coefficient of the mincerian regression with only the constant and a specific variable, say education, gives the gross or uncontrolled relative premium in terms of income variation.

The coefficient of a variable of a multivariate mincerian regression (that is, a log-linear equation with a constant and a series of additional variables) gives us the marginal controlled relative premium in terms of income variation. Thus, a tentative to isolate the effect of this variable from the possible correlations with the other variables considered.

Mincerian Regression: Individual earnings inequality within educational groups How much do variables explain? Firms fixed effects are key!



Demographics

Source: Rais microdata 1994 to 2015

RAIS - a matched employer-employee data set. Formal labor market through restricted-access administrative records with 33 million observations per year. Allows us to **track workers and firms over time**.

Minocerian Model and the Variance of Logs*

Gross Contribution to Income Inequality (%) – R^2 - CTE + VAR*

Per Capita Income		All Income Sources		Labor Earnings	
	Variable	**2008**	**2001**	**2008**	**2001**
1	Gender	0,0020	0,0002	0,0305	0,0122
2	Age	8,3227	7,0210	4,6073	4,1649
3	**Education**	**25,0497**	**31,3089**	**29,0560**	**33,3025**
4	Ethnicity	7,8616	10,3042	7,0688	9,4793
5	Migration	2,5821	2,3392	2,0506	2,0636
6	Geography	18,1450	21,1074	20,6631	23,1793

Net Contribution to Income Inequality (%) *

% Difference of R^2 without specific Variable wrt full regression R^2

Per Capita Income		All Income Sources		Labor Earnings	
	Variable	**2008**	**2001**	**2008**	**2001**
1	Gender	0,2046	0,0918	0,3178	0,1605
2	Age	14,3245	10,2909	5,5695	3,8033
3	**Education**	**34,2615**	**35,4399**	**35,7792**	**35,4216**

* Ex: in the case of education: $\ln w = \beta_0 + \beta_1$ Schooling + ϵ

Source: FGV Social from PNAD microdata

Difference in difference estimator

In economics, vast research is done analyzing the so-called experiments or quasi-experiments. To analyze a natural experiment it is necessary to have a control group, that is, a group that was not affected by the change, and a treatment group that was directly affected by the event of interest, both with similar characteristics. In order to study the differences between the two groups, pre and post-event data are needed for both groups. Thus, the sample is divided into four groups: the pre-change control group, the post-change control group, the pre-change treatment group, and the post-change treatment group.

The difference between the differences between the two periods for each of the groups is the difference in difference estimator, represented by the following equation:

g3 = (y2,t – y1,t) – (y2,c – y1,c)

Where each *y* represents the mean of the studied variable for each year and group, with the subscript number representing the sample period (*1* for before the change and *2* for after the change) and the letter representing the group to which the data belongs (*c* for the control group and *t* for the treatment group). *g3* is the so-called difference in difference estiator. Once the *g3* is obtained, the impact of the natural experiment on the variable to be explained is determined.

In order to study the impacts of local infrastructure policies between two groups, we need data at least two moments in time for both of them. Our sample is thus four fold. The interactive effect between the treatment group dummy (dT=1; dT=0 (control group omitted category)) and the time dummy (d2 =1; d2=0 (initial instant omitted category), which as we will see gives us the difference-in-difference estimator.

Mathematically, we can represent this difference-in-difference estimator (D-D) used from equations in discrete or continuous variables (for example, in the case of logistic regressions or mincerian-type per capita income equations):

Y = g0 + g1*d2 + g2*dT+ (D-D)*d2*dT + other controls

Falling Inequality– Higher individual income growth for low income groups from 2001 to 2009 :

Taking the variable of greatest interest, the difference-difference estimator (D in D), indicates higher income growth for traditionally excluded groups:

- Region: Northeast x Southeast → (6% when controlled)
- State - Maranhão x São Paulo → (12% controlled)
- Rural Area x Metro Region → (16% controlled)
- Females X Males → (-1% controlled) *exception
- Blacks X Whites → (4% controlled)
- Browns X Whites → (5% controlled)
- Construction X other sectors → (3% controlled)
- Illiterate/0 years x 12 + years → (40% controlled) *41% not controlled*

Source FGV Social from PNAD/IBGE microdata

Returns to education and intergenerational mobility Motivation:

- 2 stylized facts about Brazil:
  1. High returns and low levels of education
     - in contrast with most countries, Brazil experienced a reduction in the educational premium up to 2014
  2. Low intergenerational mobility and strong dependence of family background (persistence)
- Good, fresh and rare data for Brazil
  - PNAD 2014 supplement on Socio-Occupational Mobility
    - information on parents's education (1996, 1982 and 1976)
- Address Measurement error: make use of the information of who responded to the PNAD questionnaire on income and education, as a proxy for measurement error.

Research Questions:

1. What are the returns (wage premiums) to basic education in Brazil and how was their evolution? What are the econometric problems to measure them?
   - measurement error
   - omitted variables
2. How does the parents' education affect the returns and the educational level of their children?
3. How did intergenerational mobility in education evolve in Brazil?

Measurement error and attenuation bias

- In PNAD 2014, almost half of the sample responded to the questionnaires for themselves, which suggests a potential large problem often ignored in household survey analysis.

Education premium and measurement error – Base model	Own Person	Another Person
Education Premium	0.1339 (0.0026)	0.1060 (0.0035)
R-squared	0.4753	0.4081
Observations	5,871	2,536

PNAD 2014 supplement microdata.

- A key implication is the occurrence of attenuation bias in the education coefficient. greater and statistically significant in the sample of own respondents.

Selectivity and availability bias:

- 46 per cent of the males responded to the question about education for themselves, the corresponding number for the women is 65 per cent, which may well affect the education premium results.
- Standard logistic regression matching procedure in which we created two equal-sized and more comparable samples regarding the profile of the respondents;

Education premium and measurement error – matched sample	Own Person	Another Person
Education Premium	0.1200 (0.0039)	0.1053 (0.0037)
R-squared	0.4576	0.4093
Observations	2,293	2,275

PNAD 2014 supplement microdata.

- In the matched sample, the difference of the R-squared is still significant but a little bit smaller, the same happening for the years of schooling coefficient

Selectivity and availability bias in relation with parents education

- One concern is that the sample profile that responded to the questions regarding parents' education differ, This selectivity could also bias the results.

Table 4: Education premium and omitted variables - 2014 restricted sample

		Without Parents' Education	With Father's Education	With Mother's Education	Both Parents' Education	Highest Educational Level
Education Premium		0.1261 (0.0021)	0.0991 (0.0025)	0.1023 (0.0024)	0.0961 (0.0025)	0.0991 (0.0025)
Parent's Education		-	0.0435 (0.0020)	0.0402 (0.0021)	-	0.0412 (0.0020)
R-squared		0.4552	0.4858	0.4795	0.4881	0.4832
Observations		8,409	8,409	8,409	8,409	8,409

Source: Author's calculation based on PNAD microdata.

We observe a reduction in the wage premiums when we include information on the parents' background and the magnitude of the drop is bigger, when we have the education level of both parents, in this case, a reduction of 24 per cent happened.

Education Premium from 1996 to 2014

To assess the changes in the wage premiums from 1996 to 2014, we piled up the PNADs. We can estimate the coefficient as the change in education premiuns.

Changes in the educational premium from 1996 to 2014

	Without Parents' Education	With Father's Education	With Mother's Education	Both Parents' Education	Highest Educational Level
Education Premium	0.1277 *(0.0019)*	0.1110 *(0.0020)*	0.1136 *(0.0020)*	0.1090 *(0.0020)*	0.1105 *(0.0020)*
Parents Coefficient	-	0.0416 *(0.0017)*	0.0403 *(0.0018)*		
Change	-0.0018 * *(0.0026)*	-0.0117 *(0.0026)*	-0.0125 *(0.0026)*	-0.0141 *(0.0026)*	-0.0114 *(0.0026)*
R-squared	0.4940	0.5135	0.5106	0.5159	0.5122
Observations	15,912	15,912	15,912	15,912	15,912

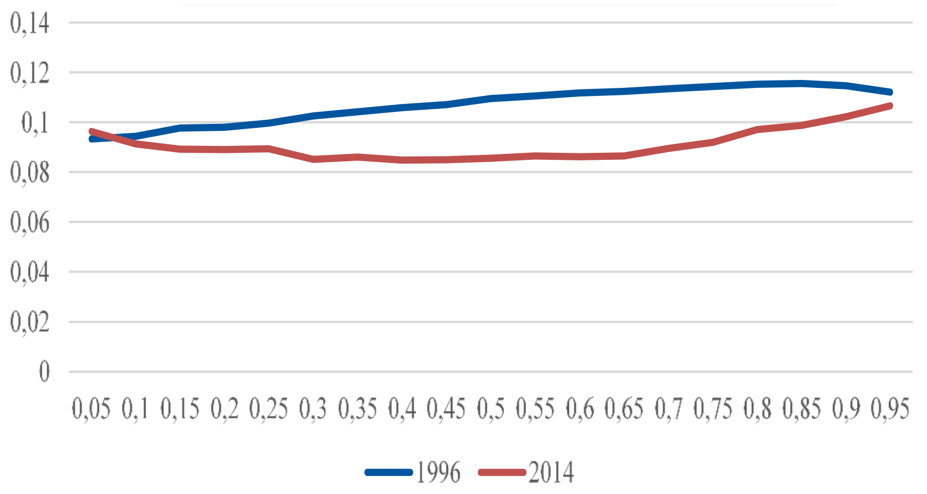
Source: Author's calculation based on PNAD microdata.

The estimates point to a reduction in the educational premium from 1996 to 2014, although the coefficient which captures this change is not statistically significant in the most basic specification without the education of the parents. However, when we include the information on the parents' educational background, the reductions in the wage premiums for the period are higher and the coefficient becomes statistically significant.

Quantile regressions

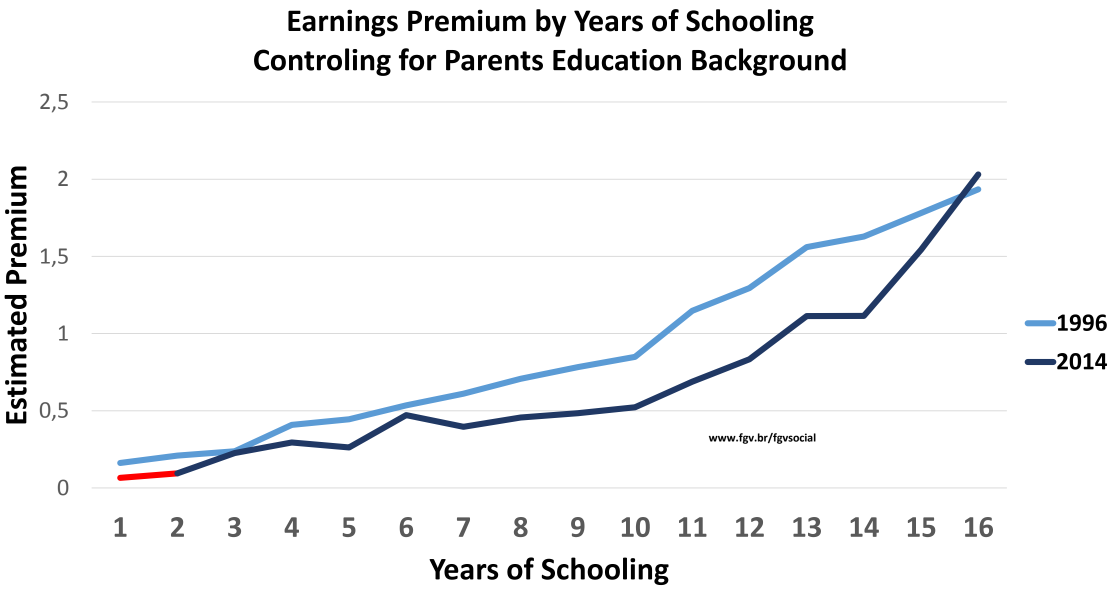
When we compare the same specification across the two different years, we find that the wage premiums are smaller in 2014 in comparison with 1996 for the entire distribution, with the exception of the first vintile. On the other hand, the reductions are smaller at the basis and at the top of the income distribution and bigger at the middle of the distribution.

Education premiums by vintiles of the income distribution – with both parents' education



Source: Author's calculation based on PNAD microdat

CHANGE IN EARNINGS SCHOOL PREMIUM 1996 to 2014



Fonte: FGV Social a partir dos microdados da PNAD 1996 e 2014 Suplemento/IBGE

Intergenerational mobility

Transition matrix for individuals with 15 to 59 years old - 2014

	Education of the Children					
	Preschool	Elementary School	Middle School	High School	Undergraduate	Graduate
Total	0.06	4.84	31.27	40.24	18.07	0.82
Education of the Father						
Preschool	2.41	6.84	32.91	33.52	14.97	0
Elementary School	0.05	5.56	30.6	42.1	17.64	0.86
Middle School	0.12	0.04	20.47	56.35	21.6	0.79
High School	0	0.2	7.25	45.47	44.25	2.24
Undergraduate	0.03	0.05	2.19	19.55	70.66	7.09
Graduate	0	0	1.32	8.27	65.96	22.75

Source: PNAD microdata.

On the top of the distribution, we have that among fathers with an undergraduate degree, approximately 70.66 per cent of their children achieved the same level and 7.09 per cent got a graduate degree. Among fathers that completed high school, 45.47 per cent achieved the same level and 44.25 percent got an undergraduate degree. Therefore, it looks like there is some upward mobility even though the persistence is still high.

Intergenerational education mobility

A simple Markovian regression model of transmission of education given by:

$$S_i = \alpha + S'_{pi}\beta + x'_i\gamma + \varepsilon_i$$

where S_i is the level of schooling of the individual i, S_{pi} is a 2x1vector with the level of schooling of the parents, β is a 2x1 vector and x_i is a vector of covariates.

	1996	2014
Persistence (Father's Education Coefficient)	**0.7045 (0.0038)**	**0.4730 (0.0058)**
R-squared	**0.3897**	**0.3974**
Observations	**92,978**	**16,284**

Intergenerational mobility

Behrman et al. (2001), Gasparini et al. (2017), Ferreira and Velloso (2003)

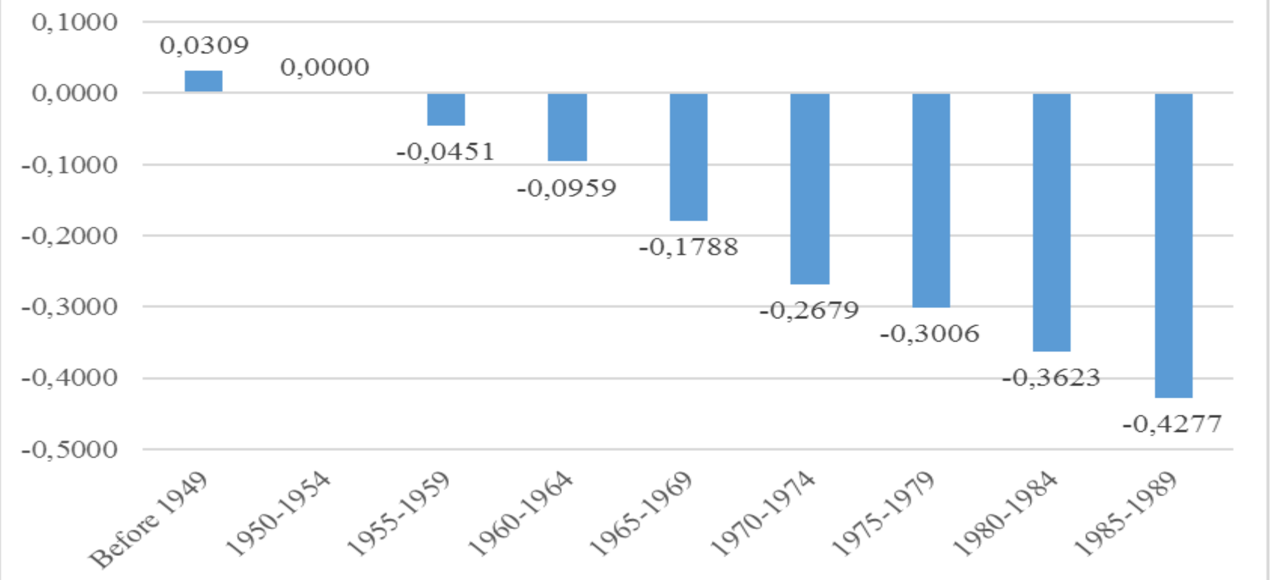
ESTUDOS DE PERSISTÊNCIA INTERGERACIONAL DE EDUCAÇÃO

Autor	Grau de persistência educacional	País
Borjas (1992)	0,25	Estados Unidos
Couch e Dunn (1997)	0,27	Estados Unidos
Mulligan (1997)	0,32	Estados Unidos
Behrman, Gaviria e Székely (2001)	0,35	Estados Unidos
Couch e Dunn (1997)	0,20	Alemanha
Behrman, Gaviria e Székely (2001)	0,70	Brasil
Behrman, Gaviria e Székely (2001)	0,70	Colômbia
Behrman, Gaviria e Székely (2001)	0,50	México
Behrman, Gaviria e Székely (2001)	0,50	Peru
Lillard e Willis (1994)	0,19	Malásia

Ferreira e Velloso 2003

How did intergenerational mobility in education evolved?

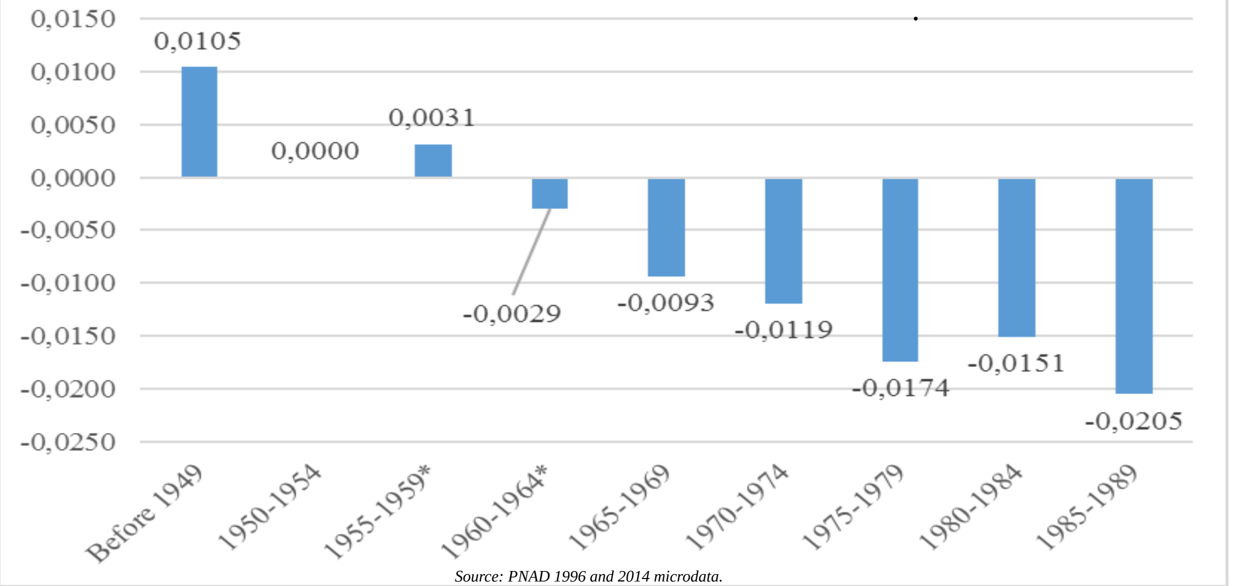
Persistence in the Intergenerational Mobility of Education by Cohorts – Interaction between fathers education and cohort effects



Source: PNAD 1996 and 2014 microdata.

What was the evolution of wage premiums with respect to schooling?

Differences in the Education Premiums by Cohorts - Interaction between individual schooling and cohort effects



Source: PNAD 1996 and 2014 microdata.

Conclusions

We used a dataset that contains family educational background with 2 objectives:

1) provide **new estimates** of the level, distribution and evolution of education premium between PNAD 1996 and 2014.

Regarding **measurement error,** the empirical strategy is to make use of the information of who responded to the PNAD questionnaire but controlling for availability biases. We find evidence of **attenuation bias** which reduces mean returns from education **between 14% and 31.5%**. **Omitting parents' education information increases the premium estimates by 24%.**

Possibility of comparing omitted bias impacts across a period of sharp earnings inequality fall observed between 1996 and 2014. **The fall of education premium turns out to be heavily underestimated when we do not take family background into account.** The highest fall of returns occurred in intermediary levels of education and income.

2) Assess **how parents' education affects the educational outcomes of their children** and how it has evolved over the last years. We find a reduction on the **intergenerational persistence of education from 0.7 to 0.47 between 1996 and 2014**.

Cohort effects regarding intergenerational mobility show that the fall in the persistence of education is also stronger for younger cohorts, coinciding with the fall of education premiums.

Abstract:

Education-related changes are often argued as the main reasons for changes in earnings distribution. However, omitted variable and measurement error biases possibly affect econometric estimates of these effects. Brazil experienced a sharp fall of individual labour income inequality between 1996 and 2014. Coincidentally, in the Brazilian National Household Sample Survey (PNAD) there are special supplements on family background in these two years that allow us to better address the role played by falling education returns. This paper takes advantage of this information to provide new estimates of the level and evolution of the returns to education in Brazil using variable premiums by education level, quantile regressions, and pseudo panels. Regarding measurement error, the empirical strategy is to make use of the information of who responded to the PNAD questionnaire but controlling for availability biases. We find evidence of attenuation bias which reduces mean returns from education between 14 and 31.5 per cent. On the other hand, omitting parents' education information also accounting for selectivity issues reduces the premium estimates by 24 per cent. Perhaps more importantly, the fall of education premium is heavily underestimated when we do not take family background into account. The highest fall of returns occurred in intermediary levels of education and income. Cohort effects also show that the reduction in the educational premium has been going on for several generations. Finally, we assess how parents' education affects the educational outcomes of their children and how the intergenerational mobility of education has evolved over the last years. We find a reduction on the intergenerational persistence of education from 0.7 to 0.47 between 1996 and 2014. Cohort effects regarding intergenerational mobility also show that the fall in the persistence of education is also stronger for younger cohorts, which coincides with the fall of education premiums.