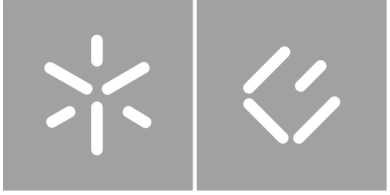
Universidade do Minho
Escola de Economia e Gestão

Rafaela Aparecida Henz Ferreira Dias

The impact of upper secondary school expansion on wages: Evidence from Brazil

July, 2020



Universidade do Minho
Escola de Economia e Gestão

Rafaela Aparecida Henz Ferreira Dias

The impact of upper secondary school expansion on wages: Evidence from Brazil

Master thesis

Master in Economics

Advisors: **Francisco José Veiga** e **João Cerejeira Silva**

July, 2020

Acknowledgments

First, I would like to thanks my parents, Viviane and Júlio, who always support my decisions. My family and friends, who patiently listened to my cries when the database was bugging. Specially Renata, Guilherme, and Laís, who spent a lot of time validating my feelings.

Special thanks to Maurício, Anita, and Maya, who made my days in DC happier and cozy.

To Caio Piza, who gave the first start in this project.

I am grateful to the DIME Analytics team for their support.

And last, but not least, I would like to thank my advisors and professors, João, Francisco, and Miguel who helped and motivated me to continue.

Thank you all, you made this work possible.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo Este trabalho investiga o impacto da expansão do sistema de ensino básico e secundário no Brasil, ocorrido nas últimas décadas, nos retornos salariais. Para tal, recorreu-se a modelos econométricos que tivessem em conta a endogeneidade das variáveis explictativas, nomeadamente o método de Variáveis Instrumentais, para estimar um modelo de tipo minceriano, usando como variável instrumental as taxas de matrículas no ensino médio tradicional.

A base de dados é um painel que reúne informações do Censo Escolar, Relação Anual de Informações Sociais, IBGE e DataSus, tem início em 1998 e termina em 2014.

Os resultados apontam para uma diminuição nos retornos para os trabalhadores com nível de ensino secundário e superior. Assim, a qualidade do sistema de ensino praticado no país é colocado em questão, uma vez que a redução dos retornos do ensino secundário indicam uma diferença não significativa face aos valores dos retornos para o ensino médio nos anos mais recentes da análise.

Palavras-chave: Retornos para educação, Brasil, Salários, Expansão escolar

Abstract

This work investigates the impact of the Brazilian basic and upper school expansion, which occurred in the last decades, on the wage´ returns. In order to estimate the Mincer equation, it uses econometric models that took into account the endogeneity of the explanatory variables, namely, the Instrumental Variables (IV). It used the enrollment rate on the traditional upper secondary as an instrument.

The database is organized as a panel that goes from 1998 to 2014. Its sources are the School Cesus, Annual Relation of Social Information, IBGE, and DataSus.

The results point into a diminishing in the returns to education at the tertiary and upper secondary levels. Because of it, the quality of education is put in check. Once the reduction on the upper secondary returns shows a non-significant differences face of the most recent years in analysis.

Keywords: Returns to education, Brazil, Wages, School Expansion

Contents

1 Introduction 10

2 Literature Review 11

2.1 The volume and the returns 12

2.2 Quality matters 13

2.3 Concepts 14

2.3.1 The return to education 14

2.3.2 Credention and Screening effects 15

3 The Brazilian education system 15

3.1 The structure 15

3.2 The upper secondary panorama 16

4 The system's expansion 17

4.1 Quality and inequalities 22

5 Methodology and data 22

5.1 Econometric tools 23

5.2 The formula 25

5.3 Our Model 26

5.4 Data sources 26

5.5 Summary Table and Variables 28

5.5.1 Main Variables 28

5.5.2 Control Variables 28

5.5.3 Base category 28

5.5.4 Instruments 29

6 Results 30

7 Robustness Checks 32

7.1 The total number of schools as an instrument 33

7.2 Tagging a new school 33

7.3 Robustness Results 33

8 Discussion 35
9 Conclusion 36
10 Appendix 39

List of Tables

1 Summary statistics 29
2 Results 30
3 Robustness 34
4 Correlation: Demography and Schools 39
5 First-Stage 39

List of Figures

1 Upper secondary school expansion index between 1998 and 2014, public and private. 17
2 Total of upper secondary school between 1998 and 2014, public and private. 18
3 Upper secondary school sector share between 1998 and 2014, public and private. 18
4 Upper secondary school sector share between 1998 and 2014, public and private by region 19
5 Upper secondary school expansion between 1998 and 2014 at rural and urban 19
6 Total number of upper secondary schools between 1998 and 2014 at rural and urban 20
7 Upper secondary school expansion between 1998 and 2014, share of rural and urban schools . . 20
8 Upper secondary school expansion between 1998 and 2014 at rural and urban areas by region . 21
9 Total of enrollments by age between 1998 and 2014 21
10 Enrollment rate mean between 1999 and 2014 22
11 Returns to education; cross-section IV regression 32

1 Introduction

Education is the foundation of societies. Without it, we could not testify the technological advances that exist today. More than that, we would not even be able to read this text. For this reason, the educational system has an important role. It instructs people, development their citizenship skills, and allows them to innovate.

For many reasons, the connection between the labor market and education has been a traditional topic for economists. It helps to understand how educational decisions can impact individuals´ choices. For example, if they should get one additional year of education or not. Considering the idea that higher levels of education lead to higher wages (Kolesnikova et al. 2010). And, that they seek to optimize their costs and benefits choice (Murnane 2013). They could get more training or not.

Another motivation to study this subject is that if the policymakers could get a better understanding of the mechanism that forge the differential in the returns to each level of formal education. They would make rational decisions more efficiently (Weale 1993). And it also helps to analyze distributional income effects and to design programs that create incentives for educational investments (Patrinos 2016).

This financing has two channels. One happens through the private sources, and in this case, the policymaker can design student loans and establish repayment rates more consistently. The second uses the public structure, in which case they can design education finance programs that help the student to pay its scholar bills at private entities (Patrinos 2016).

Considered as a key strategy for economic development, the expansion of access to education is not a guarantee that the country would have better economic conditions. The education quality between the developing and Organisation for Economic Co-operation and Development (OECD) countries are very discrepant. And this difference is growing. The scenario turns out to be worst when considering also the learning achievements. So, there is a necessity to create an educational reform through the national systems to diminish that difference. But, its success is only guaranteed if both student access and the teaching's quality would be ensured by the public power (Hanushek and Wößmann 2007).

In contrast with the developed countries that have a well established educational system, the developing countries have a growing structure of schools. The Brazilian system, in particular, has been showing an expanding performance since the 1990s. This work will focus on the causal effect of the upper secondary school expansion on returns to schooling, and also it aims at answering the following questions: What happened with the returns to education through the expansion process? How much is the premium that the people with the secondary school diploma will receive in contrast with the other that did not?

The starting hypothesis relies on the reduction costs caused by an increase in the number of local

schools. Better explaining, the creation of a school that is near to students home would diminish the transportation costs, and is expected that the attendance increases, augmenting the probability of the pupil to finish the mandatory formal education.

The inflation hypothesis relies upon a short-term growth in the returns to education, but a fall right after. So, in consequence of the labor market adjustment, the effects on the medium-term would be negative (Alda et al. 2020). Besides, there is also a screening effect hypothesis. In sum, it is the diplomas' value deterioration because it does not ensure worker literacy (Patrinos 2016).

The changes in the return to education also bring up to light some questions about the relevance of the education quality in a context of a heterogeneous country, and how this could produce effects on the wage distribution along the time. Ferreira et al. (2008) point out that the reduction of the returns to school between the school-level groups is one of the explanations to the inequality reduction in the country between 1993 and 2004.

This thesis starts with a literature review about the returns to education, and some important concepts. The third chapter presents the Brazilian education panorama. The fourth explores the education expansion. The model and the data are going to be explained in the fifth chapter. The sixth chapter presents the results. In the seventh, we proceed with a robustness check. Following by the discussion. The last chapter will conclude.

2 Literature Review

The literature about returns to education is vast, and it can rely on different themes such as universal education, compulsory school, quality of teaching, and others. To keep the focus on the upper secondary expansion, the literature was selected adopting the criteria of primary relevance, provenance, and related to education expansion. So, this literature review is organized by theme and categories.

The first theme "Volume and the returns" considered only the expansion in volume and the returns to education, using as categories the macro studies, worldwide study cases with expansion and Brazilian studies.

The second theme "Quality matters" starts explaining more further the link between the education and the returns, and than focusing on the macroeconomic analysis to finish with the country case.

The following theme "Concepts" seeks to clarify the concepts that will be used along with the text, given the literature interpretation.

2.1 The volume and the returns

Macroeconomic studies that compare multiples countries are important. Because it is possible to identify some tendencies of the returns to schooling. Patrinos (2016) and Psacharopoulos and Patrinos (2018) show that the prevailing private rate of return to an additional year of schooling is about 9%. Which can be higher to developed countries and in Sub-Saharan Africa, but it tends to decline modestly with time. The main conclusion in the first work focuses on the association between education and earnings at an individual level. It shows that schooling can create an augment in earnings. The second reveals that the returns to primary education have a downward path. Due to the higher social profitability in obtaining basic education, the performance in low-income countries is better. The private return, on the other hand, decreased at all levels of education, and also depends on the per capita countries' income.

Duflo in her works (Duflo 2001, 2004) studied the effect of the education expansion in Indonesia. The main difference between this expansion to our case relies on the program design. The Indonesian's program happened in two different points in time and was directed specifically to children, while the Brazilian expansion studied in this thesis was directed to teenagers and also had a continuous development. She validated the hypothesis of a positive linkage between the construction of a new school and wages, in a short-run analysis. Although, when it was studied on the medium-run the effect was negative because the attainment to school grew faster than wages.

Alda et al. (2020) connect the premium wages and educational expansion in West Germany. Focusing on the medium and high skilled labor, they found that education expansion has a negative effect on those who have a technological certificate but do not have the propaedeutic diploma. That it is aimed to give the student a pass to tertiary education. In contrast, that did not occur with the workers that have the tertiary educational level. This study did not find any empirical evidence of educational inflation. So, they can not confirm that the linkage between an increase in the supply of credentials and the negative effect on rates of return.

It is important to remember that the educational structure in Germany differs from the study case. For this reason, the paper does not address the upper secondary school. Although it reveals that the vocational occupations with low skills tend to get worsen than unskilled professionals. And, the high-skilled workers tend to be stable. When compared to people that have only the lower secondary school, an individual with the unfinished upper secondary school would get worse than those who finished (Alda et al. 2020).

Using the OLS and IV method, Štefánik and Horvát (2015) analyze the tertiary education expansion in Austria, Germany, Czech Republic, and Poland. They found that after 2009 the returns before and after the expansion were the same. But, when considering just Austria and Germany, the result was different.

The returns after the expansion were higher. This highlights the country particular characteristics have an important function at the individual outcomes.

Besides, they show a theoretical panorama about the screening hypothesis. Summarizing, it works as a filter to classify individuals by qualities, so an increase in access to education would weaker the filter, and the diploma value would fall. The big conclusion is that the country-specific factors play a role in the final consequences that the tertiary expansion has on societies and individuals (Štefánik and Horvát 2015).

On the other hand, the literature which focuses on the quantity expansion in the Brazilian education system and its relation with the returns to education, launch some light on the inflation hypothesis. First, Arnold and Jalles (2014) and Ferreira et al. (2016) show that the returns in Brazil had a falling tendency. Once the supply of labor with certificates and diplomas have grown in consequence of the expansion program, the returns to upper-secondary and tertiary education decrease. In particular, the first study detected that this reduction has a magnitude of around 4%. In concordance with this tendency Wang (2019) analyzed the quality cohort in Brazil given the expansion of the tertiary education, putting in evidence that the tertiary premium becomes smaller to younger cohorts, that was affected by the expansion.

While the literature education in Brazil supports the falling of the rate of returns, as shown in the macro-studies, it fails to demonstrate empirically how the expansion affected it, as have been done for Indonesia and West Germany. This work target to fulfill this demand, and contribute to the understanding of the use of the upper secondary education as an instrument to estimate the returns.

2.2 Quality matters

As aforementioned, the expansion of the educational system when it is just an increase in the number of attainments to school, and the quality of teaching is not ensured, it can mask the idea that the augment on population educational level increased (Hanushek and Wößmann 2007). To sum up, the credential value will decrease along the time, and the certificate that was obtained would not guarantee that the qualifications and skills were learned by an employee.

Hanushek and Wößmann (2007) compared the returns resulted from a quality education improvement and a quantity school increase. They detected, from a macroeconomic perspective, a positive relationship between an increase in quality and the income returns at developing countries whose implemented policies on the enhancement of the institutions, reforms of the educational structures. These can include for example the creation of strong students' performance accountability or an environment that incentives the institutions' autonomy and consequently making it possible to professors lead with local issues. In contrast, the second return type is frequently associated with a diminishing of the returns to schooling,

but this adverse behavior can be reversed if both policies of improvement and increasing were applied simultaneously, boosting the school attainment and the cognitive skills.

The literature about the relationship between Brazil's income inequality and the quality of its educational system attempts to use the data from the Brazilian National Household Survey (PNAD[1]) (Medeiros et al. 2019; Tavares and Menezes-Filho 2011), or a cross-sectional evaluation with Brazilian Census (Brotherhood et al. 2019). This work on the other hand uses an alternative source, that compiles more data about the formal labor market and the national educational system.

In the work of Brotherhood et al. (2019) they seek to identify the differences between the education quality given the heterogeneous Brazilian context. To find a good measure, they use the data from migrants at the city of São Paulo that probably have their education attended at other states to support the evidence that the return to education has a quality educational measure. Their finds are following the previous work and show that this diversification on the level of investment in different areas of a country can be reflected in an inequality income inside the national area.

Similarly, Tavares and Menezes-Filho (2011) also investigate the association between education and national income inequality. They found a rapid change of labor force educational composition. While the means return to education decreased significantly between 1995 and 2009. The returns to experience tended to be higher for more educated workers. But for the upper-secondary labor force during the time it tends to be flatter in comparison with tertiary education. Their main found was that the composition of the labor force can diminish the inequality between the educational level groups, but also can cause an increase within-group.

Medeiros et al. (2019) also includes that the attainment is influenced by region factors, such as the demographic and social background. More than that, through simulations they test various scenarios, like the increasing number of credentials, declining and rising return to education, and also with demographic inertia. They explore the argument that the increase on education investment could not result in better conditions, and even the simulation not being realistic enough it could put in evidence the limitations of that asset can have on the inequality and the poverty reduction, noticing that it could be also translated in a worse scenario.

2.3 Concepts

2.3.1 The return to education

Before the presentation of the empirical model is important to clarify some concepts that could appear along with the text, to avoid some misunderstandings.

[1] *Pesquisa Nacional por Amostra de Domicílios* in portuguese.

The first concept to be clarified is the return to education. Following Card (1999) the return to education is a random variable that may differ between the individuals, once they experience different backgrounds, such as family structure, ability, level of schooling, and others. Although there is another way to face this concept and relies on the idea that like any other investment the rate of return to education this rate is a summary of costs and benefits of an investment in education in different points of time, and can be used to explain the individual behavior in face of educational choices, such as to invest in more education or drop out school (Murnane 2013). So the individuals choose to spend more time and money to get greater lifetime wealth in return (Oreopoulos and Salvanes 2009). The third approach to this concept is to consider the return as an indicator of productivity that can be mathematically expressed as a function of the developments in market conditions, which assumes that the wage reflects the productive characteristics of the economy and workers (Alves et al. 2010; Patrinos 2016; Psacharopoulos and Patrinos 2018).

2.3.2 Credention and Screening effects

It is necessary to put the focus on another two effects. Namely the credential and the screening effects. Both of them are related to the return to education as long as the higher level of education is the source of them. The first one relies on common sense the higher education is associate with higher earning because they certify that the worker is likely to be more productive. By the other hand, the second is built on the idea that employers will hire workers with more education to reduce the risk of hiring someone with low capacities of learning (Patrinos 2016).

3 The Brazilian education system

3.1 The structure

The education system in Brazil since your beginning had changed considerably. Nowadays, it can be described by two different angles. The first one uses the national structure. While the second uses the UNSD [2] International Standard Classification of Education to distinguish each degree. This classification, particularly, allows us to do a comparison between national educational systems. Its concepts and definitions are agreed between the UNESCO members states (UNESCO 2012).

Adopting the national classification, formal education can be described as an education system which is controlled by public or recognized private entities. They have the responsability to institutionalized and organized it (UNESCO 2012). It has four levels: Early child, Elementary, High School and Higher Education. The first three are designated as basic education and are compulsory to the children from 4 to 17 years old

[2] United Nations Statistical Division

(Brasil 2013; De Castro and Tiezzi 2004; Krawczyk et al. 2009).

The beginning of education path is marked by the infant education, which is constituted by the nursery and the pre-school. Starting at age 0, it is compulsory for kids at the age of 4, when they must start in the pre-school. It has the purpose of preparing the student to the next stage. It lasts 9 years with the main goal to develop the basic skills to practice citizenship. The basic education final step is the high school. This aims to prepare the student to the adult life. Particularly, it gives them the tools to proceed to the labor market or to follow to higher education. In other words, it is a propaedeutic education (Brasil 2013; De Castro and Tiezzi 2004; Krawczyk et al. 2009).

The basic education analogous in ISCED classification is a four-stage path. It starts at level 020, named as the pre-primary school. Which can be characterized by the improvement of children's' skills. For example, social living, talking, use of logic, and reasoning. The second stage is the primary school, or level 1. It aims to provide basic numeracy and literacy skills, improve some core areas. Such as personal and social development. ISCED level 2, also called as lower secondary education, has a subject-oriented curriculum. At this stage, theoretical concepts across a diversified range of subjects are introduced. Finally, there is the ISCED level 3, which is the upper secondary education. Usually, it requires teachers who are more qualified. Because this stage offers a diverse and specialized instruction (UNESCO 2012).

From now on, along with this work, the ISCED terminology is going to be adopted. In that way, is expected to turn the future comparisons between findings easier.

3.2 The upper secondary panorama

Before the 1930s the upper secondary school in Brazil held an elitist and a sexist venue. In other words, only rich men could access formal education. With the presidency of Getúlio Vargas, the structure changed. Starting to look more similar to what stands today. With a duration of 2 years long, it was considered as a link between the lower secondary and tertiary stages. In the following decade, it had two main branches, the classic and the scientific. The first had a propadeutic character, while the second was more technical directed. Because of the economic growth in the 1960s, the need for skilled labor increased. In consequence, the demand for schools also augmented (Ribeiro 1993).

In the 1990s the Brazilian educational system had several changes. Leading it to an upper secondary expansion. Basic education becomes universal and the enrollment on traditional upper secondary school increased. On another hand, the night school enrollment rate decreased. In 1998, the first national exam was implemented (De Castro and Tiezzi 2004; Krawczyk et al. 2009; Oliveira 2004).

The technological innovations in the following decade required workers with better qualifications.

Although, the education system was still presenting an excessive number of dropping-outs, and repetition. To sum up, some students dropped the school, with the eminence of failing the scholar year, to return on the next year. So reform was implemented, focusing on put the knowledge in context. This gives the system more flexibility and diversity. The inclusion of the upper secondary as basic education, and also a final stage, grounded the curriculum guidelines. That was directed to the development of competences, abilities, and contents. It was constituted as basic preparation for work although it was not a strictly professional education (De Castro and Tiezzi 2004; Krawczyk et al. 2009; Oliveira 2004).

4 The system's expansion

The Brazilian School Census is an annual survey that gets data about the professors, students, and institutions. Given the advance of the Information and communications technology (ICT), made it possible to collect data with more precision and detail, mainly in a country with a large territorial area. In contrast, when the method of information collection changes we may have some adjustment period when the schools adopted the new system. Between the years of 2006 and 2007, the survey system changed, and for this reason, there are some unexpected breaks.

The study using the School Census is going to start with a brief analysis of the public and private sectors. The graph below takes 1998 as a year base, and it shows the index evolution of the expansion in both. It is possible to conclude at first sight that it had grown in the two. At the end of 2014, the number of public schools was 80% higher than when our series begging, while the private expansion corresponded to almost 70%. Since 2010 there is a deceleration of this process on both. But, the public evolution seems to have faster and stable progress. To refine the picture, the examination of the total of schools and the share of each one in the education market.

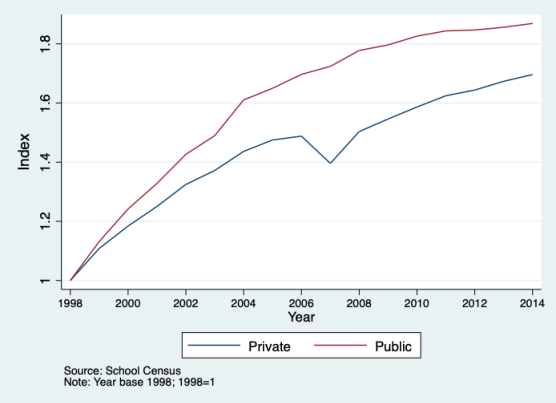


Figure 1: Upper secondary school expansion index between 1998 and 2014, public and private.

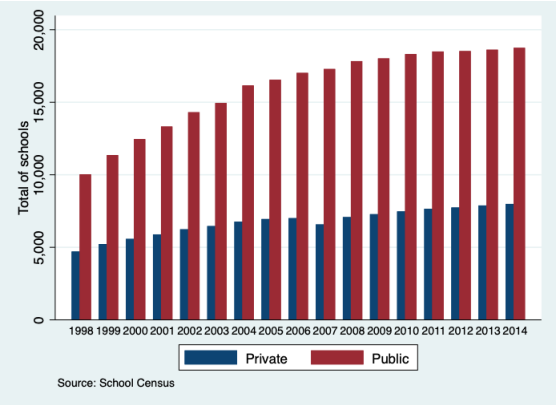


Figure 2: Total of upper secondary school between 1998 and 2014, public and private.

The graph above displays the total number of schools at the public and private sectors, red and blue respectively. It enlightens the fact that the public sector was the main engine at the increase of the supply in volume. The private sector, on the other hand, did not have the same amplification, because its initial number was smaller.

Even though, when the study focuses on the market share there was not a large change. The private market had kept a share of around 30% along with all the expansion. That evidences that the expansion was driven by both entities.

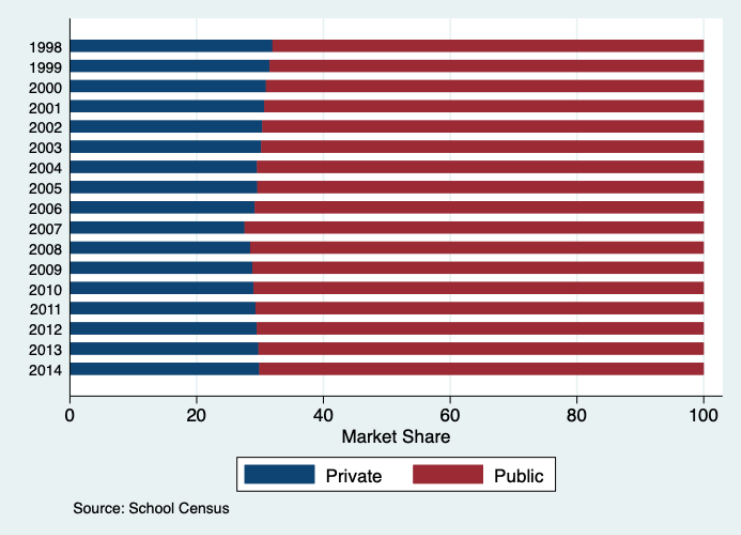


Figure 3: Upper secondary school sector share between 1998 and 2014, public and private.

The next figure displays the expansion evolution. Grouped by the five Brazilian regions: North, Northeast, Southeast, South, Central-West. The heterogeneous scenario became clear in this picture. The expansion seems to be centralized in the Southeast area, following by the Northeast and South. In contrast, the North and the central-west shows a weak development.

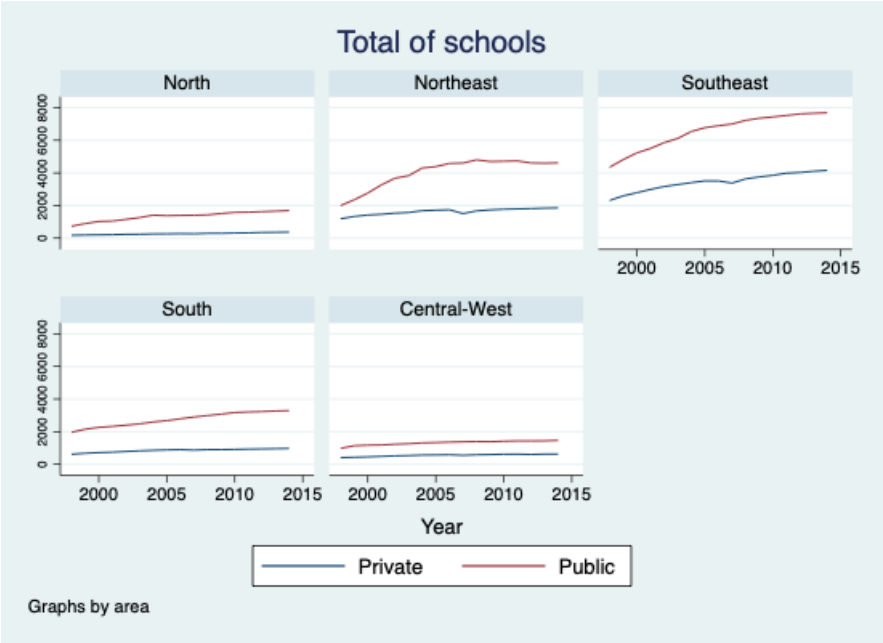


Figure 4: Upper secondary school sector share between 1998 and 2014, public and private by region

Furthermore, other aspects deserve attention. The distinct circumstances between rural and urban education. These two areas have their particularities in many aspects, economic, social, and others. According to the graph below, the rural area in Brazil changed dramatically, with fast growth in education supply in the area. In 2014 the number of schools in rural areas was more than 6 times higher.

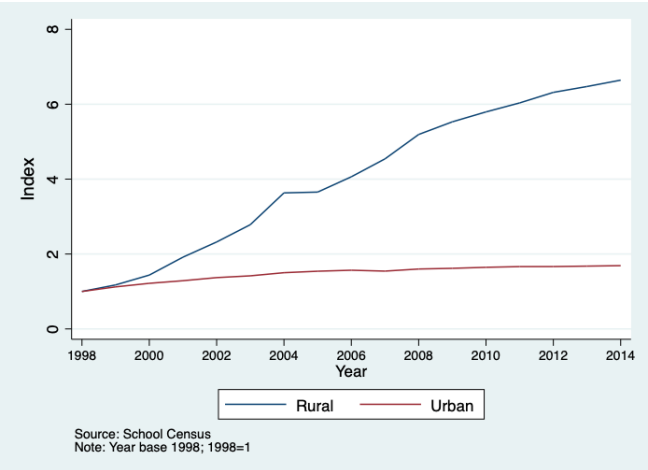


Figure 5: Upper secondary school expansion between 1998 and 2014 at rural and urban

Although, when looking at the total number the scenario changes. More than 90% of the schools are in

urban areas.

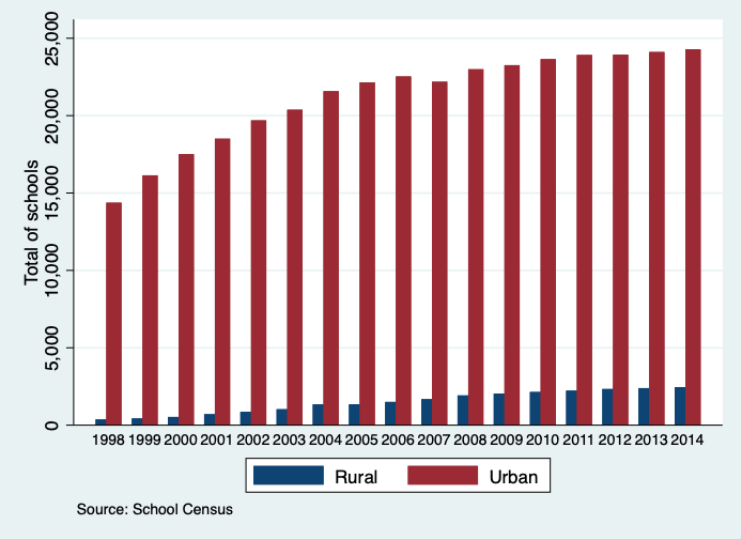


Figure 6: Total number of upper secondary schools between 1998 and 2014 at rural and urban

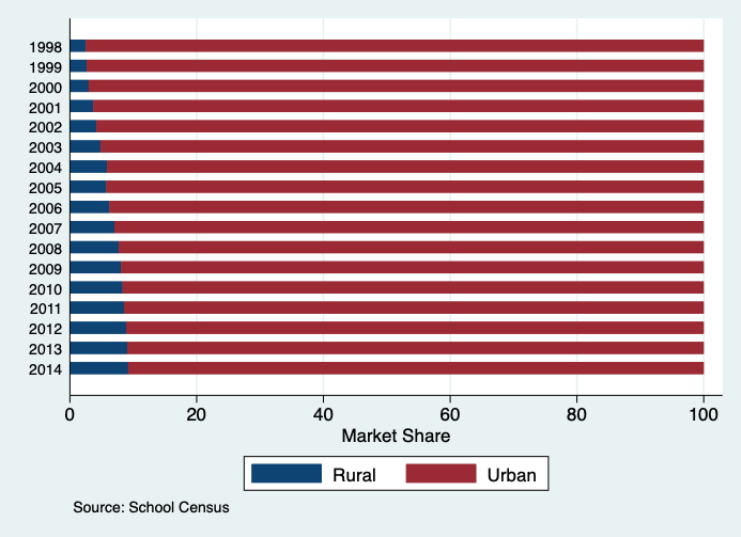


Figure 7: Upper secondary school expansion between 1998 and 2014, share of rural and urban schools

The Southeast is the region that concentrates most of the schools. Being strongly characterized by the urban school. In the Northeast, the urban curve looks stable, and the rural area presents a slight growth path. In the other regions, the evolution kept stable.

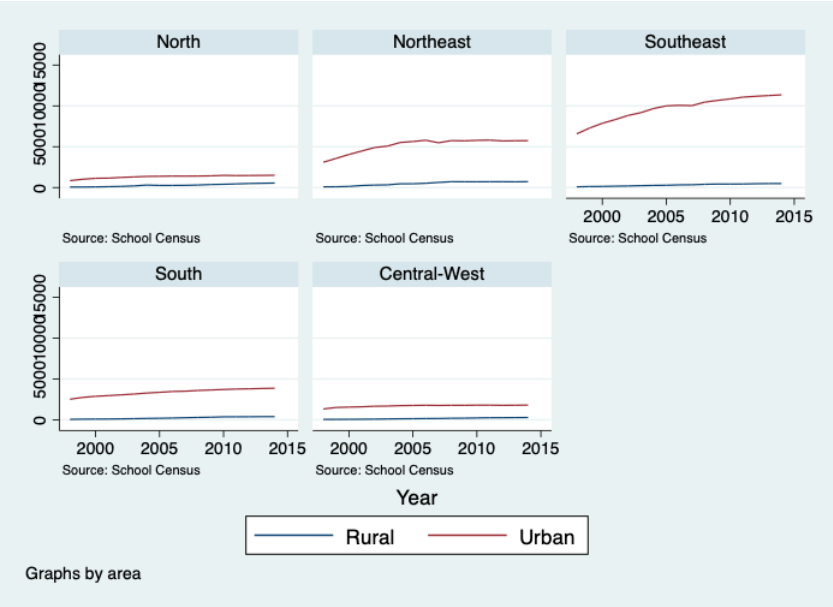


Figure 8: Upper secondary school expansion between 1998 and 2014 at rural and urban areas by region

Along the period in focus, there were some demographic changes in school structure. Observing in the following graph, the number is students that have more than 18 years old decrease along the time. While the number of students with ages between 15 and 17 increased. And, the number of students with ages between 18 and 19 lowered. This could be explained by a reduction by the school drop-out rate.

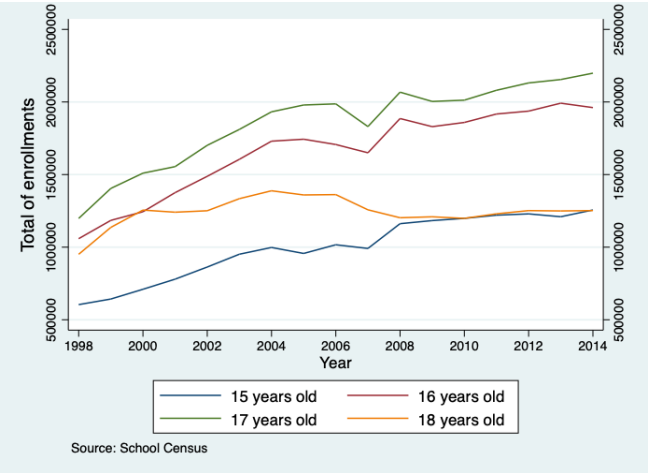


Figure 9: Total of enrollments by age between 1998 and 2014

Following the previous graph, the mean enrollment rates on upper secondary school for the students with 18 years old remain stable, while for the other ages, the data shows a growing tendency. The economic crisis in 2008 had produced many dropouts in 2009. It is not possible to ignore that the enrollment rates are still low on the big picture, and the economic panorama, as well as your diversity and differences between regions, has to take in consideration.

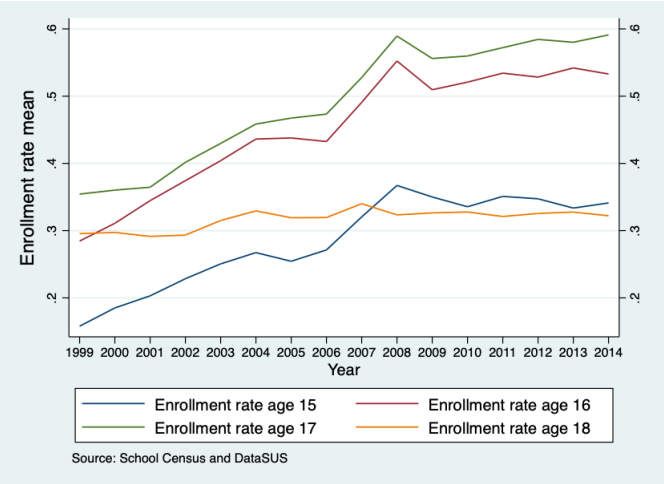


Figure 10: Enrollment rate mean between 1999 and 2014

4.1 Quality and inequalities

However, is important to keep in mind that these processes were not uniform all over the country, and some points have to be highlighted. First of all, there is an intrinsic culture of expansion in Brazilian education, which means that instead of focusing in improve quality of the system, the educational administrative body believes that is a growing demand, and consequently this creates a space for expansion without quality. So more years of poor education have the potential to generate weak academic and labor skills, and for this reason, the unchecked expansion of the educational system can be worse than expected from the conventional economics (Oliveira 2004).

First, it will produce different returns to the same level of qualification in the different regions of the country, namely the North and Northeast would be the most affected by this policy given the lack of investment in the educational system quality in that area. In contrast, the highest returns will be concentrated on the rich areas, South, and Southeast, that invest more in education (Brotherhood et al. 2019).

5 Methodology and data

If we have access to a database much more complex, with other information, then it could be possible to design an ideal experiment. That means if we have data about the individual's characteristics, their families, grades, and decisions. Such as if they drop-out the school, or not, and why. More than that, if we have access to labor market data, the education system, and also could track the migration along the time. It would be possible to separate who had access to a new school in their municipalities by whom did not receive it. This would provide us a perfect measure of the causality that the new school has on the

individual income. Another option is to apply randomized trials. Selecting municipalities to receive a new upper secondary school. Creating two statistically identical groups. This enables us to cut the selection bias, that emerges from the unique background of each municipality. And finally, achieve the true value of the causal effect of the school expansion on wages.

Instead of this perfect world, with perfect access to information, and unlimited sources. The currently available information relies on the education structure, provided by Scholar Census. There is also data with unidentified observations about the individuals' income at formal firms. Due to these restrictions, it is necessary to make some adjustments in the perfect experiment model. That focuses on quantifying the differentials in wages. Specifically, between the individuals who have the upper secondary diploma and the others that do not finish it yet.

Although, the decision of increasing the offer of upper secondary schools was not random. The correlation matrix, presented on the Table 4 in the Appendix, displays a strong, and positive, correlation between the youth demography and the schools' supply. Such a factor also influences the choice to create or close a school. As a consequence, the municipalities with larger density would have more probability to expand the schools' offer. This also can be identified in the previous section. The graphs picture that non-randomized scenario.

5.1 Econometric tools

Because there is not a randomized trial, there are other mechanisms that we can use to almost eliminate the selection bias. Specifically, the *regression*. Regressions are very common in the economic empirical field because it aimed to reduce the selection bias. It uses the characteristics which can be observed to fulfill some information gap. One key concept that we have to keep in mind, to understand the regression technique is the *ceteris paribus*. Which means "keep all the other relevant factors constant" (Angrist and Pischke 2009, 2015).

The Ordinary Least Squares, following Verbeek, is an algebraic tool. Through the minimization of the residuals, it seeks to find the best linear approach of the observational data. This tool imposes some assumptions, also called Gauss-Markov assumptions. It states that all residuals must have the same variance, and their expected value has to be zero. In this way, it ensures the homoskedasticity and avoids the existence of autocorrelation between the error terms. Another strong assumption is the independence between the error term and the other dependent variables. If all these proved to be valid, the model would produce an unbiased estimator (Verbeek 2008).

But, if the error term averages were different from zero, and if there is no independence between it

with the other explanatory variables, then the estimator will be biased. (Verbeek 2008). Because there is much-unobserved information, once it is not possible to ensure that the error term is not correlated with the error term. In this study case, for example, the unity of observation is large, and some relations between them can not be observed. For example, the population migration, that will end with the independence between the observations and between the respective error term (Wooldridge 2016). In consequence, the estimator of this model will present auto-correlation and the independence between the explanatory variable and the error term will disappear.

In a panel data scenario, it is possible to apply the fixed-effects methodology, which assumes that the unobserved individual characteristics will emerge through a dummy variable employed to distinguish it. But it also will emerge through a time dummy, that identifies the period variable. In sum, the individual characteristics that do not change through time will be captured by the dummy component, and the time component will capture that time characteristics. But, it is possible that few unobserved characteristics remain on the model and that producing bias (Angrist and Pischke 2009).

To minimize the bias we can apply the Instrumental Variables (IV) method. Briefly, it uses an auxiliary equation containing variables that capture the previous bias. Those variables are called instrumental variables. They and have to be correlated with our interest variable, but uncorrelated with the error term. If such variables exist, the model is aplicable (Angrist and Pischke 2009). It has to follow three main requirements. First, it has to affect the variable which is trying to capture the effect. Second is the independence assumption. It means that the instrument has to be independent of the other omitted variables. The third assumption is the exclusion restriction. So, the instrument only affect the dependent variable through a single channel (Angrist and Pischke 2015).

The estimation under the IV method is executed on two regression stages. The first stage is when the instrumental variable is applied and the coefficients retrieved. In the second stage, the results recorded are employed in place of the previous into the original regression. Because of it, the estimation process is denominated two stages least-squares (2SLS) (Angrist and Pischke 2009).

Because of the panel's structure, it is very hard to ensure the independence between the observations, and between them with the time frame. Consequently, there will be a serial correlation problem. So to correct it, and ensure the asymptotic validation, we clustered the model at a municipality level. Once, we have a lot of municipalities and periods the problem that is originated by the few clusters problem is also avoided. As a result, the standard errors have less chance to be biased and produce inaccurate results (Angrist and Pischke 2009).

5.2 The formula

The human capital earnings function (HCEF) was designed by Mincer in 1974 (Mincer 1974). His model is most employed to study the relationship between wages and education, on empirical studies (Patrinos 2016). Even in your most simple form, it is capable of explains around 20-35% using the observable data. Although, when it is also too parsimonious and it cannot characterize all the distribution of age, earnings, and schooling. But its simplicity allows us to use the model as a starting point (Card 1999).

Its flexibility allows researchers to establish the relationship between earning, education, and experience. It also accepts many other variables. Which turns it in a very convenient model (Psacharopoulos and Patrinos 2018). However, it has limitations. First, because it is possible to include different variables, it is difficult to compare studies. So the models are rarely the same. Second, surveys sometimes may not reflect the real mean of the population. That is, the survey-data is susceptible to a range of upward and downward biases (Patrinos 2016; Weale 1993).

The general model has the specification of the equation 1. Where y is the earning, S is the schooling level, X is the experience, and α_1 is the rate of private returns. The use of the log transformation is popular. Besides, it is close to the normal distribution and has a convenient interpretation (Card 1999). Wooldridge (2016) emphasizes that the transformation of the linear model produces a semi-elasticity of the wages for education. (Wooldridge 2016).

$$logy = \alpha_0 + \alpha_1 S + \alpha_2 X + \alpha_3 X^2 + \epsilon \tag{1}$$

The result of this estimation is the private returns to schooling (Patrinos 2016). So, it takes into account the costs of the foregone earnings and the fees or incidental expenses. And the benefits are counted as how much extra earning the individual will earn after one more education year. Although, there is the social rate, which reflects the social spending on education and social benefits. This aspect is very important, in particular to the developing countries (Weale 1993). That rate can be estimated by adopting the full-discounting method. (Psacharopoulos and Patrinos 2018).

The formula has also implications. One of them implies the unit of time employed to measure earnings can bias the outcomes. Assuming that individuals with higher levels of schooling tend to work more. The returns will be greater when using weekly or annual earnings than hourly earning. The next limitation is to admit that the returns to experience are equal for all educational levels. So if the model is too simplistic, and other observable variables are not included, there is a chance to be biased (Card 1999; Patrinos 2016).

5.3 Our Model

In this respect, and to diminish that bias, this work proposes to use control variables and the instrumental variables approach to the inference the returns to education. This technique relies on proposing an observable variable that co-variate with the schooling choices but is independent of the ability factor (Card 1999). The education expansion measured by the enrollments rates in Brazil can be used as an instrument, specifically, because we assume that the expansion of upper secondary course is independent of the experience, and the other control variables, but on the other hand it could cause an increase of the population share that has the diploma because the creation of school offer will decrease the cost to attend the school, and increase the probability of the municipality individuals to receive the diploma.

Based on the equation 1 and the previous explanation, the empirical models applied in our estimation are the following:

Mincer equation:

$$lny^a_{m,t} \quad = \quad \beta_0 \; + \; \beta_1 \widehat{TE^a_{m,t}} \; + \; \beta_1 \widehat{USE^a_{m,t}} \; + \; \boldsymbol{\gamma \Pi^a_{m,t}} \; + \; \theta^a_m \; + \; \lambda^a_t \; + \; \epsilon^a_{m,t} \quad (2)$$

Excluded instruments:

$$TE^a_{m,t} = Enroll^{a-\Delta}_{t-\Delta} + \boldsymbol{\nu \Pi^a_{m,t}} + \mu, a - \Delta = \{15, 16, 17, 18\} \quad (3)$$

$$USE^a_{m,t} = Enroll^{a-\Delta}_{t-\Delta} + \boldsymbol{\tau \Pi^a_{m,t}} + \mu, a - \Delta = \{15, 16, 17, 18\} \quad (4)$$

At equation (2) ln_y is the natural logarithm of the earnings/hour mean in the municipality m, in time t, to individuals with age a. $\mathbf{\Pi}$ are the component of the control variables, while θ^a_m is fixed effects to the age-municipality, and λ^a_t is the age- effect. ϵ is the residual. Because of the lack of information, the mean experience is not applied at our model.

In the (3) $TE^a_{m,t}$ is the share of formal workers that hold a tertiary education diploma at time t, with age a in the municipality m. While in the (4)$USE^a_{m,t}$ is the share of formal workers that hold an upper secondary diploma. At both equations $Enroll^{a-\Delta}_{t-\Delta}$ is the enrollment rate when the individual had $a - \Delta$ years old. μ is the residual.

5.4 Data sources

The data comes from different sources. Namely, the School Census, the Annual Relation of Social Information (RAIS), IBGE, and DataSUS. All them are open data and currently online.

The Brazilian Scholar Census is an annual survey that gets data about the professors, students, and institutions. The sample only kept the schools that offered the traditional upper secondary and were working in the survey year. The data collection system changed from 2006 to 2007. This change improves the information collection efficiency. It also means an increase in information quality. But some considerations have to be done. First, there is a period when the survey changes its platform, leading to an adjustment time that created an information loss. Second, the variables definition and the basic education structure changed. So the variables harmonization process may lead to inconsistencies between studies using the same database.

The Annual Relation of Social Information (RAIS) database compiles the information about the employees in the formal sector. That survey must have been answered by the employers at the end of each year. So, it has national coverage and shows information about the market formal sector. When compared with other databases, such as Pesquisa Nacional a Domicilios (PNAD), it does not show many disparities. And also offer much more information about the Brazillian formal labor market (De Negri et al. 2001; Saboia and Tolipan 1985).

Our unity of observation is the municipalities. At the moment that this thesis has been written, there are 5,570 municipalities. Its number can change with time because it is a political division. The identification system also changed, from 6 digits to 7 digits. So we link the different codes using a Table created by DataSUS. It compiled the information about municipalities' creation. After the merging process, our database computed 5558 municipalities.

As shown in the previous chapter, the changes in the educational structure were not homogeneous. In our sample, there are only the schools that were active in the year of the census and had positive enrollments. So there are years with no information about the upper secondary school in that municipality, leading to drop in all the year observation. There is some lack of data in RAIS too. We compiled the data for workers with the ages from 15 to 30, and that was working at the year before. So if one of them is missing, all the observation about it is dropped. Because of these, the panel is unbalanced.

This structure has some benefits since it compiles observational characteristics that change in time to the same units. But on the other hand, it is harder to be obtained, because it is necessary a lot of work collecting information (Wooldridge 2016).

To summarize, the panel is collapsed by the municipality, year, and worker ages. Because it is unbalanced, there are years with fewer municipalities than others. Or, if there is no worker with 15 years old in a specific municipality, it will show fewer observations. In this case, is not possible to make assumptions about it.

(Wooldridge 2016).

5.5 Summary Table and Variables

The data used in this work are summarized in Table 1. The first part corresponds to the main variables. The second is the control variables, and the third is the instrumental variables. In average each municipality had a share of 11,5% of workers that hold a tertiary diploma, and 43% holds an upper secondary diploma.

5.5.1 Main Variables

The main variables are from RAIS and retrieve the information about the educational labor shares. To be able to compare the values, the information about the shares with upper secondary incomplete or lower were omitted.

Wages

The wages are measure at local current prices. the sampling process dropped the null values. The information about the weekly hours is available. So, the arithmetical transformation to income per hour and the logarithmic change were applied.

5.5.2 Control Variables

As presented in chapter 5.1 they are observational data that can the bias between the dependent and independent variables. As the main variables, these are also from RAIS.

Females

In this context, the share of females is used to capture the bias that could emerge from the inequalities in the wage system between sex.

Firm characteristics

Continuing, the share of different sizes and sector´s firms are used to capture the bias that is caused by the frim background. That is, we assume that big firms can support high salaries, so if the municipality had more small firms, the wages would be smaller.

As well as the size, the sectors also have different capacities to support the wage costs.

CNPJ and CEI

The CNPJ and CEI are the initials for two types of the registry. The first is the firm registry, while the second is for an individual that has an economic behavior as a firm.

5.5.3 Base category

Because the data present a multilinear problem, it is necessary to omit variables to avoid it an ensure a good base of comparison. The variables omitted are:

- The share of workers that had upper secondary school incomplete or lower education.
- Firms with less than 5 employees
- Public Entities

5.5.4 Instruments

In contrast with the previous variables, the data used are from dataSUS and School Census.

Past enrollments

The enrollment rate (ER) is the ratio between the total of enrollments of individuals at age a, and a total of the resident population. The information about the enrollments is from the School Census and the estimates of the population from DataSUS.

$$Enroll_{t-\Delta}^{a-\Delta} = \frac{\text{Total of enrollments}_{m,t-\delta}^{a-\delta}}{\text{Total of the resident population}_{m,t-\delta}^{a-\delta}}$$

Because of the upper secondary schools lack, at some municipalities, it is possible to this value assumes the value of zero and bigger than one.

Table 1: Summary statistics

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
ln_y	1.176	(0.558)	-1.238	4.966	991647
TE	0.115	(0.138)	0	1	991647
USE	0.432	(0.231)	0	1	991647
% of females	0.406	(0.199)	0	1	991647
% workers between 5 and 9	0.117	(0.128)	0	1	991647
% workers between 10 and 19	0.101	(0.118)	0	1	991647
% workers between 20 and 49	0.105	(0.129)	0	1	991647
% workers between 50 and 99	0.075	(0.14)	0	1	991647
% workers between 100 and 249	0.164	(0.248)	0	1	991647
% workers between 250 and 499	0.12	(0.226)	0	1	991647
% workers between 500 and 999	0.076	(0.184)	0	1	991647
% workers $>$ = 1000	0.05	(0.146)	0	1	991647
% Private entities	0.585	(0.315)	0	1	991647
% International Organizations	0	(0.002)	0	0.5	991647
% Independent worker	0.082	(0.144)	0	1	991647
% NGOs	0.028	(0.067)	0	1	991647
% Of workers with cnpj	0.868	(0.16)	0	1	991647
Year	2006.721	(4.558)	1999	2014	991647
Past enrollment rate when 15	0.231	(0.184)	0	3.929	369178
Past enrollment rate when 16	0.387	(0.238)	0	5.214	429570
Past enrollment rate when 17	0.433	(0.223)	0	5.333	490467
Past enrollment rate when 18	0.31	(0.128)	0	4.333	551792

6 Results

The Table 2 displays the main results. It shows the estimation obtained through the OLS method at column (1) and column (2), the fixed-effect method at column (3), and the Instrumental Variables at column (4). The last, instrumented the tertiary and upper secondary education. Its first-stage is reported in the Appendix. The control variables are the same in all models. The standard errors are recorded in parenthesis.

The return to tertiary education (TE) is positive, and statistically significant, along with all the estimations. Its values increased between column (1) and (2), which means that it was being biased by the control variables. But, it starts to decrease when the model became more complex. Especially, when comparing columns (3) and (4). Since the variable was instrumented and the value reduced, the expansion affected negatively the TE. This is reasonable once with more people having access to tertiary education. It will also increase the chances to them get this certificate, because of the screening and credential effects. This, in this case, will also be intensified because of the same effect on the upper secondary sector. Once it starts to push the values down, the above levels would see its returns diminished as well.

On another hand, upper secondary education (USE) is positive and statistically significant at columns (2) and (3). Although its values decreased between the two models. It puts in evidence that the local characteristics and the time has an impact on its value. In contrast, the estimator becomes negative and lost its significance at (4). Meaning that is not possible to reject the null hypothesis. So, the USE estimator could be zero, or the value obtained is far from its real value.

Table 2: Results

	(1)	(2)	(3)	(4)
	OLS_compact	**OLS_long**	**FE**	**IV**
Variables	ln_y	ln_y	ln_y	ln_y
TE	0.782***	0.937***	0.543***	0.507**
	(0.0152)	(0.0145)	(0.00897)	(0.237)
USE	0.0112	0.142***	0.125***	-0.140
	(0.00727)	(0.00700)	(0.00433)	(0.100)
% of females		-0.0339***	-0.0915***	-0.0613***
		(0.00584)	(0.00325)	(0.0115)
% of worker between 5 and 9		0.138***	0.0654***	0.0406***

		(0.00688)	(0.00530)	(0.00690)
% of worker between 10 and 19		0.224***	0.0678***	0.0372***
		(0.00925)	(0.00602)	(0.00814)
% of worker between 20 and 49		0.322***	0.116***	0.0650***
		(0.0113)	(0.00782)	(0.0116)
% of worker between 50 and 99		0.501***	0.177***	0.0858***
		(0.0134)	(0.00911)	(0.0149)
% of worker between 100 and 249		0.513***	0.169***	0.115***
		(0.0116)	(0.00810)	(0.0170)
% of worker between 250 and 499		0.481***	0.184***	0.125***
		(0.0123)	(0.00901)	(0.0175)
% of worker between 500 and 999		0.502***	0.209***	0.132***
		(0.0124)	(0.00965)	(0.0177)
% of worker >= 1000		0.597***	0.246***	0.142***
		(0.0162)	(0.0116)	(0.0207)
%Private entities		0.329***	0.0550***	0.000694
		(0.0101)	(0.00807)	(0.0298)
%International Organizations		0.0680	0.0846	0.205
		(0.195)	(0.187)	(0.131)
%Indepedet woker		0.306***	0.0940***	0.00235
		(0.0184)	(0.0104)	(0.0366)
%NGOs		0.472***	-0.00515	-0.0756***
		(0.0260)	(0.0159)	(0.0186)
% of workers with cnpj		-0.101***	0.181***	0.109***
		(0.0150)	(0.00684)	(0.00908)
Observations	991,647	991,647	991,647	337,335
R-squared	0.688	0.722	0.835	0.754
Hansen J statistic				0.651
p-value of Hansen J statistic				0.722
First Stage F-statistic				19.23
Number of id			66,680	54,333

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

The goodness-of-fit measured by R-squared shows that the model can explain around 72% of the wage variation. Where 68% can be explained by the variation at education level variables. So, the share of the population´s educational level has a significant influence on the municipality wages. Taking into account the screening effect, in the municipalities where the number of educated people is bigger, would pay more.

One of the problems that can emerge from the IV method, is the presence of a weak instrument. To sum up, it means that the instrumental variables used are not strongly correlated with the endogenous variable. This is going to result in a biased estimator. Moreover, it will not be an asymptotic approximation. To evaluate the instruments used at column (4), we proceed the F-test and the Hansen J-statistic test (Verbeek 2008).

The F-test consists of the examination of the instruments' f-statistics. The literature(Angrist and Pischke 2009; Murteira et al. 2016; Verbeek 2008) states that if it is bigger than 10, then the instruments are in a safety zone. And can be considered a strong instrument (Angrist and Pischke 2009; Murteira et al. 2016). As we can observe at Table 2, the instruments f-statistic is bigger than 10, so they are strong instruments (Baum et al. 2003; Verbeek 2008).

To test if the instruments are valid or invalid, the Hansen J-statistics were computed. And do not reject the model specification. So the instruments applied are valid (Baum et al. 2003; Verbeek 2008).

Theses tests reassure the findings at (4) for the TE. So, using the IV method, into a cross-sectional model to retrieve the returns to education each year. The graph below displays the evolution by year, from 2003 to 2014, of the estimators. Both returns rates decreased substantially during this period.

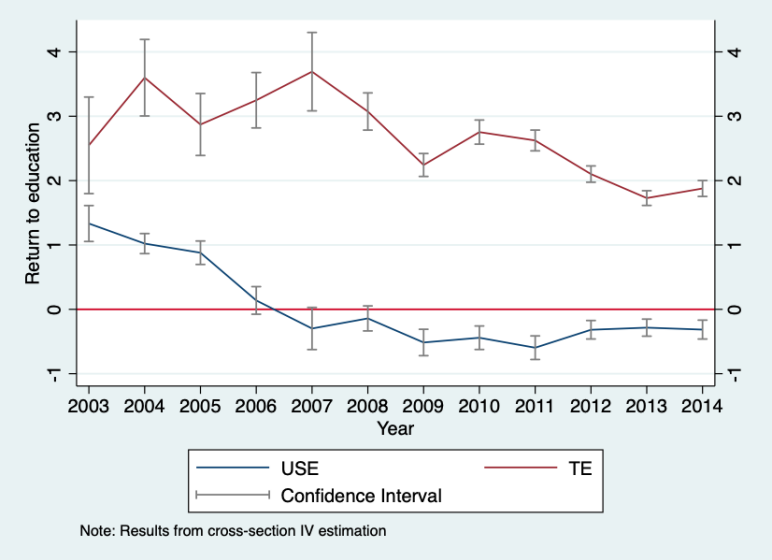


Figure 11: Returns to education; cross-section IV regression

After 2006, the estimators to USE become negative, stable, and very close to zero. In contrast, the TE presented a growth behavior in the beginning, in 2007 started to decrease. It could be resulted from the fast growth in the number of schools in the previous years and at the enrollment rates.

In figures 9 and 10 in the panorama section, the number of students stabilized after 2007. The drop in the return rates, in this case, can be seen as a result of the credential inflation in the labor market.

7 Robustness Checks

To attest to the results presented in the last chapter, we proceed with a robustness s. That will respond to the issue of the expansion measure. In particular, there are other valid ways to measure education expansion in Brazil? These measures can be instruments? The outputs, in these cases, are valid? To answer

those questions, first, we identified other measures to school expansion, using the available database. Then, we test them as instruments.

7.1 The total number of schools as an instrument

The first option for the enrollment rate instrument is the total number of schools. In brief, this variable captures the sum of active schools in each municipality. This includes public and private schools. As well as had been done with the previous instrument, this will be lagged back to when the working class was adolescents.

The variable advantage is being a measure of the supply, and as seen in the panorama section, it had increased along the time. However, this will not guarantee that the number of students attending school will increase. Moreover, sometimes schools stop offering upper secondary education from one to another. Resulting in a decrease in the school offer, an increase in the transportation costs to students, that will stop attending it.

So, this variable can be a great candidate because it reflects the expansion. But, it will not ensure an augment on the qualified labor. This weak link can put this variable in check.

7.2 Tagging a new school

Another alternative is to tag the municipality when the number of schools increases in that year. Specifically is a dummy variable that receives the value 1 if the number of schools in that municipality increased, and 0 otherwise.

The advantage of using a dummy variable is the possibility to tag only the municipality that had offer unless one new course of upper secondary school. But on the other hand, if the expansion in terms of municipalities were not too expressive, the effect of the expansion can be dissipated.

7.3 Robustness Results

The following table presents the results for the robustness checks. At columns (1), (2), and (3) we use different specifications of the model, using the total of upper secondary schools as instruments. At the column (1) the USE variable was omitted and added up at (2). At column (3), the instrument specification also includes the enrollment rate.

The columns (4), (5) and (6) are correspondent to the upper secondary school tagging used as an instrument. The same strategy was applied to them. The column (4) the USE variable was omitted and added up at column (5), and at column (6) the enrollment rate was included as an instrument.

Table 3: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	IV2	**IV3**	**IV4**	**IV5**	**IV6**	**IV7**
VARIABLES	ln_y	ln_y	ln_y	ln_y	ln_y	ln_y
TE	0.321*	1.288	0.430*	1.024	0.826	0.513**
	(0.191)	(2.303)	(0.229)	(1.218)	(1.371)	(0.234)
USE		1.885***	0.000812		-0.295	-0.143
		(0.531)	(0.0951)		(0.653)	(0.0991)
% of females	-0.0617***	-0.258**	-0.0678***	-0.100	-0.0671	-0.0615***
	(0.0112)	(0.103)	(0.0111)	(0.0672)	(0.104)	(0.0115)
% of worker between 5 and 9	0.0460***	0.102***	0.0455***	0.0427***	0.0341*	0.0405***
	(0.00579)	(0.0284)	(0.00671)	(0.00816)	(0.0207)	(0.00687)
% of worker between 10 and 19	0.0437***	0.191***	0.0457***	0.0558**	0.0320	0.0372***
	(0.00665)	(0.0329)	(0.00787)	(0.0219)	(0.0579)	(0.00813)
% of worker between 20 and 49	0.0766***	0.317***	0.0794***	0.0939***	0.0550	0.0649***
	(0.00791)	(0.0481)	(0.0111)	(0.0306)	(0.0922)	(0.0116)
% of worker between 50 and 99	0.103***	0.411***	0.105***	0.119***	0.0701	0.0857***
	(0.00839)	(0.0598)	(0.0142)	(0.0301)	(0.114)	(0.0148)
% of worker between 100 and 249	0.136***	0.459***	0.137***	0.146***	0.0946	0.115***
	(0.00793)	(0.0708)	(0.0162)	(0.0194)	(0.116)	(0.0169)
% of worker between 250 and 499	0.145***	0.488***	0.147***	0.161***	0.106	0.125***
	(0.00955)	(0.0682)	(0.0167)	(0.0294)	(0.126)	(0.0174)
% of worker between 500 and 999	0.149***	0.524***	0.154***	0.179***	0.118	0.132***
	(0.0122)	(0.0772)	(0.0169)	(0.0531)	(0.147)	(0.0177)
% of worker >= 1000	0.162***	0.596***	0.167***	0.198***	0.127	0.142***
	(0.0146)	(0.0914)	(0.0198)	(0.0637)	(0.171)	(0.0206)
%Private entities	-0.00312	0.447	0.0128	0.0987	0.0214	0.00121
	(0.0288)	(0.279)	(0.0288)	(0.177)	(0.256)	(0.0296)
%International Organizations	0.183	0.733	0.210*	0.355	0.257	0.206
	(0.123)	(0.547)	(0.125)	(0.322)	(0.407)	(0.131)
%Independet woker	0.0278	0.899***	0.0448	0.135	-0.00829	0.00246
	(0.0305)	(0.257)	(0.0349)	(0.187)	(0.378)	(0.0366)
%NGOs	-0.0662***	0.255***	-0.0599***	-0.0266	-0.0795	-0.0755***
	(0.0166)	(0.0983)	(0.0179)	(0.0699)	(0.140)	(0.0187)
% of workers with cnpj	0.110***	0.0262	0.107***	0.0884**	0.103*	0.108***
	(0.00866)	(0.0645)	(0.00873)	(0.0387)	(0.0528)	(0.00904)
Observations	337,335	337,335	337,335	337,335	337,335	337,335
R-squared	0.766	0.081	0.766	0.728	0.701	0.753
Number of of id	54,333	54,333	54,333	54,333	54,333	54,333
Hansen J-statistic	2.624	1.142	41.12	0.713	0.446	1.348
p-value of Hansen J-statistic	0.453	0.565	0.000	0.870	0.800	0.969
First stage F-statistic	43.35	0.160	9.870	0.326	0.326	9.791

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3 displays that the total number of schools is an invalid instrument. The results from column (3)

show that the interaction of that instrument with the enrollment rate creates a bias. It is expected since the increase in one is not completely independent of the other. Because it is possible to increase the enrollment rate without an increase in the number of schools. In this case, the classes will be bigger. Another scenario is the creation of a new school at the neighbor municipality, inducing a reduction of the enrollment on the first.

On the other hand, the column (6) the estimation using the dummy variable and the enrollment rate produces a valid instrument. While the (4) and (5) reveals that this instrument, when used alone, is weak and invalid. So, it will diminish the enrollment rate instrument power, but not enough to turn it invalid.

An interesting result emerges from column (6). It supports the finding of the previous section about the TE. The estimators are close to each other and with the same signal.

8 Discussion

In the previous chapters, the different aspects of the Brazilian educational system were presented. Considered as a mechanism to diminish inequality and improve the national economic environment. The policies adopted to improve access to the classroom could generate good results, but also, risks. The last one, it is not frequently debated by the government.

The results obtained in this work are under what was detected by Arnold and Jalles (2014) and Ferreira et al. (2016). There is a falling tendency on both levels investigated. In particular, for the tertiary returns, the work Wang (2019) also supports our conclusions.

The inflation hypothesis, discussed in the literature review, states that a certification increase will deteriorate the rates of return, in both educational sectors. We did not present data about the number of students that have completed the upper secondary stage. But the increase in the enrollment rate is also an indication that the dropping out, and the repetition rates decreased. Resulting in augment of the probability of obtaining the degree.

Furthermore, the robustness checks verified the instruments' alternatives to measure the school expansion. It indicates that the expansion must be quantitatively reflected on the increase of enrollments. This means that if the new school cannot attract students there will not be an impact. Once the variable fails to reflect the teaching quality offered, the model can not capture if there were an improvement or not. So the education quality evolution cannot be analyzed.

The debates, presented in chapters 2 and 3, highlight the need for policy combination to not produce damages on the labor market. The main reason is the deterioration of the credential effect. Once the wages paid to upper secondary labor are not high enough, and the returns are decreasing. And, the one more

educational year it is faced as an investment, the labor market will disincentives the students to proceed. So the damages will be spread, not only in the educational sector, but also in others, such as science, and health.

Exploring more the inequality in Brazil´s educational scenario. The expansion was not uniform in all nation surface. The southeast concentrates most of the schools. Two of the reasons are, for sure, the population density and the predominance of urban areas. The rural schools in 2014 were less than 10 %. This means that every 100 schools, only 9 were rural. The education is mostly managed by the government, unless 70% of the national education underlies its administration.

That very heterogeneous reality and what was found by Medeiros et al. (2019) points out to a need to invest in the system quality. Because there is one area that accumulates education investments, it is required policies. That could support the other regions to amortize the impact on the labor market, created by the local diploma inflation. Concomitantly there are cohort differences that would turn the scenario more difficult. But if it was complemented with a quality increase, it could generate great economic advances.

The lack of statistical significance of the USE returns puts in check the system's quality. Take into account the return as a market reflection of productivity. If we can't reject that the value can be zero. So, the education offered is incapable to be translated into a significant skill increase.

9 Conclusion

The literature review, about the education expansion and the returns to education, revealed a falling tendency on its values, globally and locally. It also indicates that an expansion when is used to reduce the income inequalities could have pros and cons. First, because it is a need to ensure the quality system. Second, it has to be reflected in the augment in the enrollments. And both, quality and quantity, channels are symbiotic. If the quality will not be ensured, it could harm the labor market and create other consequences.

Brazil has its education system centered in the southeast and urban areas. The increase in the number of schools and enrollments affected the formal labor market. Using the mincer equation, we estimated the returns to education using a panel data, that collected information about the formal labors and firms, population size, and schools' characteristics.

We use the enrollment rates as an instrument in the iv model, and it confirmed the falling tendency on the rates. And it also exposes the necessity of a more accurate analysis of the system quality. Besides, it looks that the evolution in the number of schools had stabilized. So an improvement in the quality of the already existent structure has the potential to generate good results, and revert the returns falling path.

References

Alda, H., Friedrich, A., and Rohrbach-Schmidt, D. (2020). Educational expansion and educational wage premiums of young male employees: A long-term analysis for west germany 1976–2010. *Social Science Research*, 85:102351.

Alves, N., Centeno, M., and Novo, Á. (2010). Investment in education in portugal: Returns and heterogeneity. *www. bportugal. pt Publications*, page 8.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Angrist, J. D. and Pischke, J.-S. (2015). *Mastering'Metrics: The path from cause to effect*. Princeton University Press.

Arnold, J. and Jalles, J. (2014). Dividing the pie in brazil: Income distribution, social policies and the new middle class. (1105).

Baum, C. F., Schaffer, M. E., and Stillman, S. (2003). Instrumental variables and gmm: Estimation and testing. *The Stata Journal*, 3(1):1–31.

Brasil (2013). Lei n° 12.796, de 4 de abril de 2013. altera a lei n° 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional, para dispor sobre a formação dos profissionais da educação e dar outras providências. *Diário Oficial da União*.

Brotherhood, L. M., Ferreira, P. C., and Santos, C. (2019). Education quality and returns to schooling: evidence from migrants in brazil. *Economic Development and Cultural Change*, 67(3):439–459.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, volume 3, pages 1801–1863. Elsevier.

De Castro, M. and Tiezzi, S. (2004). The reform of secondary education and the implementation of enem in brazil. In *c. Brock and S. Schwartzman (eds.), The Challenges of Education in Brazil, Symposium Books, oxford*.

De Negri, J. A., Castro, P. F. d., Souza, N. R. d., and Arbache, J. S. (2001). Mercado formal de trabalho: Comparaçao entre os microdados da rais e da pnad.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91(4):795–813.

Duflo, E. (2004). The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. *Journal of Development Economics*, 74(1):163–197.

Ferreira, F., Firpo, S. P., and Messina, J. (2016). Understanding recent dynamics of earnings inequality in brazil. *New order and progress: development and democracy in Brazil*, 187.

Ferreira, F., Leite, P., and Litchfield, J. (2008). The rise and fall of brazilian inequality: 1981–2004. *Macroeconomic Dynamics*, 12:199–230.

Hanushek, E. A. and Wößmann, L. (2007). Education quality and economic growth. Report 39511, The World Bank.

Kolesnikova, N. et al. (2010). The return to education isn't calculated easily. *The Regional Economist*, 18(1):12–13.

Krawczyk, N. et al. (2009). O ensino médio no brasil.

Medeiros, M., Barbosa, R. J., and Carvalhaes, F. (2019). Educational expansion, inequality and poverty reduction in brazil: A simulation study. *Research in Social Stratification and Mobility*, page 100458.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Murnane, R. J. (2013). Us high school graduation rates: Patterns and explanations. *Journal of Economic Literature*, 51(2):370–422.

Murteira, J., Castro, V., and Martins, R. (2016). Introdução à econometria. *Coimbra: Almedina*.

Oliveira, J. B. (2004). *Expansion and inequality in Brazilian education*.

Oreopoulos, P. and Salvanes, K. G. (2009). How large are returns to schooling? hint: Money isn't everything. Technical report, National Bureau of Economic Research.

Patrinos, H. A. (2016). Estimating the return to schooling using the mincer equation. *IZAWorld of Labor*.

Psacharopoulos, G. and Patrinos, H. A. (2018). Returns to investment in education.

Ribeiro, P. R. M. (1993). História da educação escolar no brasil: notas para uma reflexão. *Paidéia (Ribeirão Preto)*, (4):15–30.

Saboia, J. L. and Tolipan, R. M. (1985). A relação anual de informações sociais (rais) e o mercado formal de trabalho no brasil: uma nota.

Štefánik, M. and Horvát, P. (2015). Is tertiary education expansion observable in private returns to education?(evidence for middle and eastern european countries). *European Educational Research Journal*, 14(5):418–429.

Tavares, P. A. and Menezes-Filho, N. A. (2011). Human capital and the recent fall of earnings inequality in brazil. *Brazilian review of econometrics*, 31(2):231–257.

UNESCO (2012). *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics.

Verbeek, M. (2008). *A guide to modern econometrics*. John Wiley & Sons.

Wang, Y. (2019). Education expansion and decline in cohort quality in brazil: 1995–2013. *Applied Economics Letters*, 26(7):613–617.

Weale, M. (1993). A critical evaluation of rate of return analysis. *The Economic Journal*, 103(418):729–737.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

10 Appendix

Table 4: Correlation: Demography and Schools

	ΔSchool	Young Demography	Total Upper Sec. School
ΔSchool	1.0000		
Young Demography	0.0940	1.0000	
Total Upper Sec. School	0.0964	0.9765	1.0000

Table 5: First-Stage

	(1) TE	(2) USE
Past enrollment rate when 15	0.0170***	0.0142
	(3.41)	(1.89)
Past enrollment rate when 16	0.00777	0.0296***
	(1.73)	(4.51)
Past enrollment rate when 17	0.00754	0.0161*
	(1.82)	(2.44)
Past enrollment rate when 18	-0.0122*	0.0290***
	(-2.55)	(3.91)
% of females	0.0547***	0.0755***
	(19.43)	(17.42)
% of worker between 5 and 9	0.00459	-0.0322***
	(1.18)	(-4.94)
% of worker between 10 and 19	-0.0172***	-0.0687***
	(-4.52)	(-9.43)
% of worker between 20 and 49	-0.0245***	-0.114***
	(-6.20)	(-15.06)
% of worker between 50 and 99	-0.0241***	-0.151***
	(-5.55)	(-17.35)
% of worker between 100 and 249	-0.0146**	-0.164***
	(-3.25)	(-18.62)
% of worker between 250 and 499	-0.0231***	-0.170***
	(-4.81)	(-18.98)
% of worker between 500 and 999	-0.0429***	-0.176***
	(-8.22)	(-18.16)
% of worker >= 1000	-0.0517***	-0.203***
	(-9.42)	(-18.21)
%Private entities	-0.145***	-0.165***
	(-33.71)	(-20.15)
%International Organizations	-0.241***	-0.169
	(-6.17)	(-1.52)
%Indepenedet woker	-0.153***	-0.382***
	(-23.49)	(-29.53)
%NGOs	-0.0567***	-0.142***

	(-5.41)	(-9.04)
% of workers with cnpj	0.0316***	0.0289**
	(7.77)	(3.21)
year==2000	-	-
	(.)	(.)
year==2001	-	-
	(.)	(.)
year==2002	-	-
	(.)	(.)
year==2003	-0.0233***	-0.229***
	(-12.78)	(-58.55)
year==2004	-0.0214***	-0.201***
	(-13.31)	(-60.68)
year==2005	-0.0202***	-0.171***
	(-13.64)	(-59.36)
year==2006	-0.0222***	-0.146***
	(-16.29)	(-56.10)
year==2007	-0.0183***	-0.124***
	(-14.52)	(-53.28)
year==2008	-0.0176***	-0.102***
	(-15.20)	(-46.75)
year==2009	-0.0140***	-0.0863***
	(-11.91)	(-41.71)
year==2010	-0.0122***	-0.0642***
	(-11.77)	(-37.25)
year==2011	-0.0101***	-0.0467***
	(-11.01)	(-31.19)
year==2012	-0.0107***	-0.0271***
	(-13.11)	(-20.87)
year==2013	-0.00284***	-0.0164***
	(-4.18)	(-16.34)
year==2014	-	-
	(.)	(.)
Observations	337335	337335

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$