Derick Ribeiro da Costa Almeida

HUMAN CAPITAL DISPARITIES AND EARNINGS INEQUALITY: A REAPPRAISAL OF THE PORTUGUESE PRIVATE LABOR MARKET

Dissertation of the Master in Economics/Economic Growth and Structural Policies, presented to the Faculty of Economics of the University of Coimbra to obtain the Master's degree

Advisors: Maria Adelaide P. Silva Duarte, PhD

João Alberto Sousa Andrade, PhD

January 2020

Acknowledgements

I'd like to start expressing my gratitude to both of the advisors of this thesis, Professor Maria Adelaide Duarte and Professor João Sousa Andrade, for helping me to better explore the vast and multidisciplinary topic under investigation and, also, for making a high priority the use of elaborate statistical tools, such as the software R, to further my capacity to deliver high quality quantative results.

I'd also like to thank Fundação de Ciência e Tecnologia (FCT) for the scholarship provided to undertake reasearch within the WISER Project Portugal, a project coordenated by Professor Marta Simões at the Center for Business and Economic Research (CeBER) based at University of Coimbra, and for the possibility of investigating such an important subject to national public policy.

A special thank you goes to my parents and my brother, loved ones that always believed in my potential and supported me, in all dimensions, to establish my roots in Portugal. And also to my group of friends here in Portugal, A Real Tertúlia (Guilherme Shirazawa, João Bento, Diogo Mazeron, Francisco Vieira and José Miguel Rebolho), for giving me the friendship and support I needed during this one and a half years of course.

Cofinanciado por:

Referência do projeto:

POCI-01-0145-FEDER-029365 | PTDC/EGE-ECO/29365/2017

WISER Portugal – Welfare Intervention by the State and Economic Resilience in Portugal

Financiamento:
Projeto cofinanciado pelo COMPETE 2020, Portugal 2020 e União Europeia, através do FEDER e pela Fundação para a Ciência e Tecnologia I.P./MCTES através de fundos nacionais (PIDDAC).

Abstract: Income inequality in Portugal, although high compared to other west European countries, has fallen in the recent decade. It is known from the economic literature for well over 50 years that income is strongly associated with education and its distribution in society. In order to understand the dynamics of these variables and how they relate to each other for the case of Portugal, the current research first provides a moving picture of the inequalities in Human Capital, proxied by years of schooling, for the Portuguese private labor market between 1987 and 2017, showing that it generally increases until 2007 and decreases from this year until 2017, an achievement of educational policies in Portugal throughout the 20th century. Through the decomposition of Generalized Entropy (GE) indices for before tax income inequality, we also observe that inequalities in Human Capital have been and still are a major factor driving inequalities in income, although less important in the last decade. In order to understand other dimensions of this reduction, we estimate private wage premiums for different levels of schooling, using an updated Mincerian earnings function with important control variables, such as regions, economic activity, gender and work position. In the end, we see that there are two important forces operating over falling earnings inequality in Portugal and they are reductions in inequalities in Human Capital and compressed returns to schooling, mainly in higher education.
Keywords: Human Capital, Portugal, Mincer, inequality, returns to schooling, decomposition
JEL Classification: D310, I24

Resumo: A desigualdade de rendimentos em Portugal, apesar de elevada em relação a outros países europeus do ocidente, tem caído na última década. É conhecido da literatura económica já há mais de 50 anos que os rendimentos estão fortemente associados à educação e à sua distribuição na sociedade. Para que possamos entender a dinâmica e relação entre essas variáveis para o caso de Portugal, esta pesquisa procura mostrar a evolução da desigualdade do Capital Humano, aqui aproximado pelos anos de escolaridade, para o mercado de trabalho privado português entre 1987 e 2017, mostrando que esta, em geral, aumenta até 2007 e reduz-se a partir de então, um feito das políticas educacionais promovidas pelo governo português desde o século XX. A partir da decomposição de Indicadores Generalizados de Entropia para a desigualdade de rendimentos antes dos impostos, nós observamos que as desigualdades de Capital Humano foram e continuam a ser um fator importante para a desigualdade de rendimentos, apesar de ter a sua importância reduzida na última década. Para que possamos entender essa redução, também estimamos o prémio salarial privado para diferentes níveis de escolaridade, utilizando uma especificação mais recente da equação de rendimentos Minceriana com importantes variáveis de controle. No final, vemos que há duas forças atuando sobre a redução recente na desigualdade de rendimentos em Portugal e elas são as reduções nas desigualdades de Capital Humano e a compressão dos retornos da escolaridade, principalmente na educação superior.

Palavras-chave: Capital Humano, Portugal, Mincer, desigualdade, retornos da escolaridade, decomposição

JEL Classification: D310, I24

Summary

Index of tables

Index of figures

1. Introduction

In the European community, Portugal is a country known for having below average achievements in education, be it in terms of early leavers from education and training, lower participation rate in tertiary education or even underachievement in some fields, such as mathematics. Nonetheless, progress has been made in this front, through government effort in stablishing minimum requirements related to years of schooling, progressively raising from 6 years in the 1960's to 12 years in the 2000`s. As it is well known in the economic literature, both micro and macroeconomic, education and how it is distributed in society is one of the most important elements related to income and income inequality, which, in the case of Portugal, has been falling in the last decade. Therefore, in order to understand better this relationship for the case of Portugal, it is important to characterize the distribution of Human Capital in the country, proxied by the years of schooling, i.e. the evolution of its main different moments. In this research, the analysis of Human Capital and earnings focus the Portuguese private labor market as whole.

The distribution features of Human Capital can reveal a lot about how disperse the years of schooling attainment is across the population at stake and around what level of schooling the sample under consideration is concentrated. At that level of analysis, the questions this research tries to answer are: i) by how much has the representative Portuguese worker increased his years of formal schooling during the last 30 years? ii) what has been the evolution pattern of inequality in years of formal schooling among Portuguese workers in the same period? iii) how strong has been the importance of Human Capital inequality in before tax income[1] inequality in the private work place? The answer to these questions coupled with the estimation of rates of private return to different levels of education in Portugal is the key to understand in a more complete and detailed way the interactions between Human Capital and income in the country.

Section 2 of this work addresses the concept of Human Capital, its measurement and its association with the distribution of income among individuals. Therefore, the current research discusses Mincerian functions that relate in a microeconomic perspective investments in education with the income individuals receive in the labor market and how an empirical version of that equation can be used to study this relationship using estimated rates of return on educational attainment.

[1] We use before tax income, income and earnings interchangeably

Moving towards the methods of analysis, section 3 introduces the dataset used to conduct the analysis as well as the variables of interest. Also in this section, there is a description of the main inequality indicators and its properties, especially the strict additive decomposability property of Generalized Entropy Indexes (GEI) as well as the econometric specification used throughout the work to analyze the behavior of rates of return on different levels of education for the years 2010 and 2015 . In this section, we also present the Oaxaca-Blinder decomposition of mean values, which is going to be an important method to understand the sources of variations of average hourly real wages between 2010 and 2015. In section 4, we introduce and discuss the results obtained by the analysis of the distribution of Human Capital and the decomposition of the Generalized Entropy (GE – Theil N) index of earnings inequality for the period 1987-2017, as well as the results from the regression using the Mincerian earnings function and the Oaxaca-Blinder decomposition. In section 5 we present the concluding remarks of the results. In section 6, the list of references is displayed and, after, the Appendix compiles a series of Mincerian regressions for specific Portuguese regions that were left aside from the main text, but, nonetheless are important for future researches, especially in improving the measures of Human Capital for Portugal.

2. Human Capital and income inequality: an overview of selected literature

2.1. Human Capital: its definition, measures and how it relates to income

The studies of Human Capital as in important component for the dynamics of economic growth, that is of income per capita, were influenced by the works of Shultz (1960) and Becker (1962,1964) that established the links between accumulation of Human Capital, in terms first of formal education, and income from a microeconomic perspective. The authors used models of constrained utility maximization to study the allocation of time between education and production of final goods in order to assess how individuals make decisions regarding how much to invest in Human Capital and the returns associated with it. On this path, Shultz (1960) went on to study other dimensions of Human Capital, particularly on the job-training, and how it is optimally accumulated by individuals.

In the economic literature, Human Capital came to be generally expressed as the years of formal schooling of the working population, but in fact it is a multidimensional concept that also involves the quality of learning, the health condition of the work force as well as informal training acquired during the production process (Savvides & Stengos, 2009).

The Human Capital concept has also been an important factor to understand how economies grow over time and not only how it is accumulated by economic agents. Exogenous growth models of great importance, such as Romer, Mankiw and Weil (1992), as well as endogenous growth models, such as Romer's (1986) and Lucas' (1988), explicitly use Human Capital as a determinant of the steady state equilibrium of income per capita in a country. Benabou (1996) and Aghion (1997) take this idea further in an AK endogenous growth set up to show that not only Human Capital is important for the dynamics of an economy but also that its distribution in society can have growth implications. Assuming market failures in the credit market, where individuals with less than the average Human Capital in the economy cannot finance education expenditures, Aghion (1997) shows that the growth rate of the economy decline with increasing inequality in education.

The relationship between Human Capital and earnings appears as an empirical testable expression with Mincer (1958,1974), adapting, on a microeconomic level, the idea of an optimizing agent that decides to allocate time between producing and investing in the accumulation of years of schooling to an empirical framework.

In his most influential working paper, Mincer (1974) states that the potential gross income of the worker in a given period of time (E_t) depends on an initial level of potential income (E_0), determined by exogenous intrinsic characteristics of the individual, an exogenous rate of return on Human Capital investments (r) and, finally, net investments in Human Capital itself (C_t).

$$E_t = E_0 + r\sum_{t=0}^{t-1} C_t \quad (1)$$

In this expression, C_t is a pecuniary measurement of investment in Human Capital, such as direct expenditures in schooling and foregone income. Because this variable is hard to obtain, Mincer (1974) suggests an alteration to consider the variable of investment in Human Capital as a ratio between direct costs and annual potential gross income. This modification allows the variable to be measured in terms of the relative time spent on the accumulation of knowledge in a given year. Assuming $K_j = \frac{C_t}{E_t}$, doing the iterations to find a solution for the difference equation of earnings and linearizing by taking logarithms, the expression becomes:

$$lnE_t = lnE_0 + r\sum_{j=0}^{t-1} K_j \quad (2)$$

The term K_j, which represents the relative amount of net investments in Human Capital, can be broken down in two major forms of investments, according to Mincer (1974): one for the schooling period (K_i) and another for the post-schooling period (K_j).

$$lnE_t = lnE_0 + r_s \sum_{i=0}^{t-1} K_i + r_p \sum_{j=0}^{t-1} K_j \quad (3)$$

Through this operation, Mincer (1974) states that there is no reason to believe that the rental rate of Human Capital should be equal to schooling and post-schooling knowledge and opens up the possibility of analyzing how on the job-training and experience can influence the individual's earnings profile.

In order to proceed to an empirical specification of the earnings function, it is important to adapt the dependent variable of the theoretical model, which is gross or potential earnings, to an approximation to an observable measure of income. Potential income, according to Mincer (1974), represents the income that the individual would receive if he stopped investing in the accumulation of Human Capital, which doesn't seem to reflect labor income statistics. Individuals in the labor force are still paying for educational costs, which requires a concept of net, rather than gross, income. Mincer (1974), therefore, equates observable income to net income. In this set up, $Y_t = E_t(1 - K_t)$ which yields $ln\, Y_t = ln\, E_t + \ln{(1 - K_t)}$.

Another aspect that has to be taken into account in this process is the empirical representation of both K_i and K_j. Mincer (1974) argues that K_i is approximately equal to 1, since the opportunity costs of and expenditures in education tend to be equal to potential income in that phase of investment, while K_j follows the hypothesis that relative investments in Human Capital decline over time due to decreased benefits of additional units of Human Capital after school. The simplest formulation for K_j is a linear function of time, according to Mincer (1974) and it takes the form: $K_j = K_0 + \frac{K_0}{T}t$, where T is the length of working life. Combining both adjustments to the initial specification gives:

$$\ln Y_t = \ln E_0 + r_s s + r_p K_o t - \frac{r_p K_0}{2T} t^2 + \ln(1 - K_t) \ (4)$$

In this final equation, a new variable s appears to represent the accumulation of investments in formal schooling and it is also possible to identify the marginal decreasing returns over time to the accumulation of on the job-training (t^2). The structure of this equation is the fundamental building block in empirical estimates of the parameters (rates of return) associated to formal education and experience. One drawback, nonetheless, is related to the fact that Mincer (1974) specification does not account for non-linearity in rates of return to school, as pointed out by Trostel (2004), and a factor that will be considered when we develop the econometric specification for our study.

Apart from the theoretical advances in the understanding of how Human Capital relates to income and its growth patterns, the empirical validation of this relationship is conditioned on appropriate ways to measure Human Capital. Given the complexity of the concept, economists have struggled to create an index that would concisely summarize all of its dimensions. Instead, economists worked to understand and measure a particular dimension of Human Capital, which is formal education. The years of schooling were the easiest step to grasp that dimension, but it soon showed to be partially inadequate, due to the heterogeneity of schooling quality across countries. The same years of schooling in different places did not necessarily represent the same amount of knowledge, which, in the end, is the real covariate of real wages. Angrist, Djankov, Goldberg & Patrinos (2019) argue that the gap between years of schooling and learning is important in today's society and more severe in developing countries. While enrollment rates have increased over time, the authors show that average learning across countries has stagnated or even decreased over time.

Therefore, in order to construct a more accurate index for the quality of education across countries, Angrist, Djankov, Goldberg & Patrinos (2019) use international standardized tests, such as the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS), and regional assessments made specifically in developing countries to create comparable data, through an "exchange rate" adjustment between these two kinds of standardized tests, and a global distribution of educational quality.

This exchange rate is created by the authors using countries that participate in both international and regional standardized tests and use that exchange rate to adjust average grades of other countries in the same region that participated only in the regional assessment into grades of global comparison. The assumptions of applying such methodology are that the sample of the population of countries that participate in both kinds of tests are similar in terms of the population they represent, the tests should assess the same kinds of proficiency in different areas of knowledge (such as math, reading and science) and the exchange rate created should account only for test differences rather than country-specific effects, according to Angrist, Djankov, Goldberg & Patrinos (2019). The authors are also well aware of the limitations of such methodology derived from applying the same exchange rate over long periods of analyses. All regions of the world are accounted in this study.

Angrist et al. (2019) come up with four interesting stylized facts regarding the measures of Human Capital adjusted for the quality of learning. The first is related to the fact that, as mentioned earlier, years of schooling and learning are not one in the same. A number of countries in the sample used by the authors show increasing enrollment rates in primary school and a stagnant or even decreasing rate of knowledge retainment. The second relates to the so-called gender gap in which, historically, male students are more numerous than female, but, with this new Human Capital indicator, it is shown that female students learn more than its counterparts, inverting the gender gap. The third one is connected to the evidence that economic growth is more associated with learning than with years of schooling. The authors conduct a series of regressions, considering different types of growth theories and econometric specifications. The last relates to heterogeneity in the association between income per capita and learning, showing that the association between these two variables change depending of the level of income per capita. By Angrist *et al.* (2019) calculations, the second and fourth quartiles of the global distribution of income are the ones where learning correlates more with growth.

Another important way of measuring Human Capital is the index developed by Mulligan and Sala-i-Martin (1994) that adjusts the years of schooling of the working force by the wage rate associated to each worker. As explained in Mulligan and Sala-i-Martin (1995), the reasoning behind this method to estimate Human Capital is to break with restricting assumptions related to the average years of schooling method developed by Barro and Lee (1993), regarding the perfect substitutability among workers with different levels of schooling, the idea that differences in productivity are proportional to the years of schooling and that the elasticity of substitution are constant among workers with different levels of schooling. According to Mulligan and Sala-i-Martin (1995), adjusting the years of schooling by the respective labor income would show that workers with the same level of schooling, but working in different areas, would have their skills awarded differently, as does Pereira (2005) for the Portuguese economy.

2.2. Human capital inequality in Portugal: static and dynamic analysis

Given the theoretical importance of Human Capital distribution for the performance of economic growth and income distribution in any country, several studies with focus on the Portuguese economy were carried out in order to unveil the main characteristics of this variable and how it relates to income distribution. Duarte, Fidalgo & Simões (2010) conducted an empirical analysis for 1986, 1996 and 2005 by computing the years of schooling of the working population selected as the Human Capital proxy by using *Quadros de Pessoal* (QP) database that compiles economic and socio-demographic information about each worker in the private sector across the country. The authors calculated a series of descriptive statistics for those periods and found out that the average level of education of the Portuguese working population increased over time: from 5.46 years initially to 6.59 and, finally, to 7.80 by the end of the period considered.

Other results are worth mention. The authors observed an average decrease in inequality levels for the distribution of years of schooling in Portugal using several inequality measures, such as the Gini coefficient, the Atkinson index with parameters $\varepsilon = 0.5$ and $\varepsilon = 2.0$ – a reference value that indicates the degree of inequality aversion of the population, with higher numbers related to higher inequality aversion – and the Theil's first measure.

The authors also test a *Kuznets hypothesis* for the distribution of Human Capital in order to identify whether or not the shape of the relationship between the average level of education and education inequality is an inverted U. Dividing the population by 20 districts, the authors compute inequality indices for each of these territorial units and use this data to perform a panel data regression on initial levels of inequality and test an inverted parabola specification. The results obtained by the authors support the Kuznets human capital hypothesis for Portugal.

Other studies, such as the one conducted by Rodrigues (1996) aims at the decomposition of adult equivalent income[2] inequality by years of schooling for the Portuguese economy considering a 10 years interval (from 1980 to 1990). The decomposition allows to estimate how much of the income inequality is due to inequality of income within education groups and between different education groups. The author uses the Theil's N measure from the class of the Generalized Entropy Indices, because of its properties of strictly additivity, which makes the intra-group inequality component unbiased to income distribution shocks within the total population. Additionally, the disaggregation of the Theil's N index enables the analysis of inequality over time covering three dimensions: within group inequality, between group inequality and population composition, according to Cowell (2011).

Dividing the dataset into categories ranging from socio-demographic to gender, Rodrigues (1996) concludes that differences in schooling levels are the main characteristic associated with adult equivalent income inequality for the time period under analysis. Applying

[2] Adult equivalent income is obtained by dividing total household income to the number of family members adjusted by a factor that gives more weight to adults, and particularly to the individual that is the income source of the family, and less weight to children.

the decomposition method to the Theil's N, the author identifies that 21.0% of total inequality comes from inequalities between education groups in 1980, a number that amounts to 27.2%, 10 years later. Another interesting insight from the decomposition of inequality during this period is that most of the static inequality comes from within-group variations in income rather than inter-group for all types of categories analyzed.

Also dealing with questions regarding the distribution of income according to levels of education, but with a different methodology, Campos and Reis (2017) calculate wage premiums for Portuguese workers using a modified Mincerian specification. In their study, that covers the period 1986-2013, the authors estimated rates of return to higher education relative to those with secondary education that rose from 32.8% to 47.7% in that period. Although, if we take a closer look between 2010 and 2013, it is already possible to observe that those returns have been declining. This behavior has implications on how important the difference in years of schooling is to the differences in income, opening up an important gateway to study possible associations of declining rates of returns on the income distribution.

3. Methodology

3.1. *Quadros de Pessoal:* a description of the data set and the adjustments required

Quadros de Pessoal (QP) is the statistical source underlying this research. It is an annual inquiry, conducted by the Portuguese Ministry of Labor, which compiles information about individual workers of the private sector. It covers a series of variables, such as sex, total amount of hours worked – divided in regular and extra hours worked -, regular income as well as income from extra hours worked before tax, sector of activity, region, levels of education, work position, age, gender and tenure.

This dataset compiles information from 1986 until 2017. The information regarding the worker's age for the period 1986 until 2009 is missing and the same occurs between 2015 and 2017 for the sector of economic activity where individual workers were producing goods and services as well as information regarding the region where these workers were working, possibly due to technical and administrative reasons. This research uses information at NUTS2 (Nomenclature of Territorial Units for Statistics) level, which divides Portugal in seven different regions, five of them laying inside the continent and the other two, the autonomous regions, are located outside the continent, namely: i) *Norte*; ii) *Centro*; iii) *Área Metropolitana de Lisboa*; iv) *Alentejo*; v) *Algarve*; vi) *Açores*; and vii) *Madeira*.

The first type of analysis uses all of the available temporal dataset to create a moving picture of the distribution and inequality of Human Capital in the Portuguese private labor market and its association with earnings inequality through the decomposition of Theil's N index. The second type of analysis uses information of 2010 and 2015 to generate estimates of the rate of return of schooling levels.

In order to implement the statistical and econometric analysis we use the R software. The first task consisted in taking off from the workable dataset containing 30 years' worth of information all individuals that received as their basic monthly income an amount less than the stipulated by the legal minimum wage. The second one mitigates the presence of outliers in the database by skipping all monthly earnings observations that fell into the top 0.5% of the distribution.

The database also suffered other adjustments. Some of them are related to unspecified information about workers, such as work position or level of schooling attained, resulting on the removal of all observations that lack that information; other adjustments regarding the conversion of categorical or nominal variables, such as the maximum level of schooling attainment, to numerical values. In the case of schooling levels, it was used a table of conversion to years of schooling taking into consideration all the changes in terms of years of schooling necessary to attain a certain level of education, especially after the Bologna agreement[3].

For the years 2010 and 2015, for which it is necessary to work with an approximation for the worker's years of experience, it was created a variable for experience based on a standard method in the earnings function literature which, as in Lemieux (2006), is:

$$Exp_i = (Age_i - 6) - S_i \quad (5)$$

Exp_i is the amount of experience of individual *i*, Age_i is the age of individual *i* and S_i stands for the maximum years of schooling of that individual. The subtraction by 6 reflects the fact that, on average, an individual starts studying at the age of six.

To obtain the monthly real wages, we multiplied total earnings, considering basic and extra payments, by a price deflator, that takes on a value of 100 for 2010. In order to calculate hourly real wages, we simply divide total earnings by total hours worked for each individual.

[3] See table A1 in the Appendix for the table of conversion

For the period 1987, 1997, 2007 and 2017, our dataset contains around 2,000,000 observations. For 2010 and 2015, the dataset used contains, respectively, 2,222,797 observations and 2,125,163 observations, considering a series of dummy matrices generated for this research.

3.2. Inequality indices

The approach to measure the evolution of years of schooling (taken here as a *proxy* of Human Capital without any adjustment by the quality of learning) and its level of inequality in the Portuguese private labor market is usual in this type of literature. For the first part, we use descriptive statistics from the distribution of human capital, such as quantiles. For the second, the measurement of inequality of years of schooling, being one of the main features of this research, has also taken on more complex and refined indicators than the variance, standard deviation or the interquartile range of the distribution being analyzed. For that purpose, many indices are to be calculated. The most used are as follows:

$$Gini = 1 + \left(\frac{1}{n}\right) - \left(\frac{2}{n^2\mu}\right)\sum_{i=1}^{n}(n - i + 1)Y_i \quad (6)$$

$$Atkinson = 1 - \left(\frac{1}{\mu}\right)\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}(y_i^{1-\varepsilon})\right]^{\frac{1}{1-\varepsilon}} \quad (7)$$

$$Theil\ T = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(\frac{y_i}{\mu}\right)\ln\left(\frac{y_i}{\mu}\right) \quad (8)$$

$$Theil\ N = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\ln\left(\frac{y_i}{\mu}\right) \quad (9)$$

In the expressions above, y_i refers to the variable of interest, while n stands for the number of individuals in the population under analysis and μ for the average of the variable of interest across individuals. All these indices satisfy important mathematical axioms which, according to Cowell (2011) are the following: i) anonymity: any reordering of the elements of a vector should not affect the value of inequality; ii) scale invariance: if all elements of the underlying distribution are multiplied by a scalar, the value of inequality is to remain the same; iii) population principle: if two or more identical populations are aggregated, the inequality index should remain unaltered; iv) transfer principle (Dalton-Pigou principle): any transfer from a richer individual to a poorer individual, on the condition that their relative position in the distribution is maintained, should decrease the inequality index.

The Gini coefficient is a measure of inequality based on the Lorenz curve, a geometric representation of a function that relates the cumulative percentage of the population with the cumulative percentage of the variable of interest – in this case Human Capital. Its main feature is the importance it accords to data dispersion around the mode of the distribution. That characteristic makes this indicator more sensible to changes in the middle of the distribution than in other parts of it. The index values range between 0 and 1, being higher values related to higher levels of inequality.

The Atkinson index is a more flexible inequality indicator due to its dependence on a social preference parameter (ε). This parameter can be imposed exogenously and regulates the degree of social aversion to inequality, penalizing the dispersion of data between the extremes of the distribution.

Theil's T and N, also known as Theil 1 and 0, fall into the category of the Generalized Entropy Indexes (GEI), which are indicators that measure the level of chaotic behavior within a distribution. Theil T is known to give more weight to inequality on the right part of the distribution while Theil N, on the left part.

Within the category of the GEI, it is devoted particular attention to the Theil's N index, which unlike other indicators can be subjected to strict additive decomposability. This feature makes possible the decomposition of income inequality into two different types of inequality, an intra-group element and a between-group component. As in Cowell (2000) it enables the analysis of inequality over time covering three dimensions: within group inequality, between group inequality and population composition

$$Theil\ N = \sum_{i=1}^{\gamma} \left(\frac{n_\gamma}{n}\right) N_\gamma + \sum_{i=1}^{\gamma} \left(\frac{n_\gamma}{n}\right) ln \left(\frac{n_\gamma}{n}\right) \quad (10)$$

The group in question can be defined according to the research's objectives. In this case we define the group by years of schooling or Human Capital level, following Rodrigues (1996), i.e. we want to know how much of earnings inequalities comes from Human Capital inequalities. The first component of the decomposition gives a weighted average of the inequality of income for each level of Human Capital, while the second one is the level of inequality for a distribution of averages conditional on the level of Human Capital. The weights used in the decomposition come from the share of the population with a specific level of schooling.

3.3. Econometric specification of the earnings function

Following the theoretical Mincerian specification outlined in section 2.1.we study the relationship between earnings and investments in formal schooling and on the job-training. The variable schooling (S) is treated here as a categorical variable, as in Campos and Reis (2017), in order to account for different private returns on education depending on the maximum level of education attained by individual i. As we mentioned before, this specification is important for our research, because it enables returns to education to vary across different levels of schooling and it offers another methodology to account for the importance of Human Capital distribution to earnings variations.

$$y_i = \alpha + \sum_{j=2}^{5} \beta_j S_{j,i} + \theta_1 Exp_i + \theta_2 (Exp_i)^2 + \theta_3 Tenure_i + \boldsymbol{Z}_{\boldsymbol{i}}^T \boldsymbol{\rho} + \varepsilon_i \quad (11)$$

In the above equation, y_i already stands for the logarithm of hourly real earnings, while $S_{j,i}$ ={1,2,3,4,5} is a dummy variable that takes the value one when individuals: 1) have less than 9 years of formal schooling; 2) have exactly 9 years of formal schooling; 3) have exactly 12 years of formal schooling; 4) have between, but not included, 12 and 15 years of formal schooling; 5) have 15 or more years of schooling. It is possible to observe that individuals with less than nine years of formal schooling are disregarded in the equation, since the introduction of it would generate collinearity problems. Therefore, this group of individuals is the reference for calculating the wage premiums of those with higher schooling levels. Exp_i is the number of years of experience calculated by the formula introduced in section 3.1., while $Tenure_i$ is the number of years the employee *i* has been working at the current company. The term $\boldsymbol{Z^T}_{\boldsymbol{i}}\boldsymbol{\rho}$ represents a dot product that takes the form:

$$\boldsymbol{Z^T}_{\boldsymbol{i}}\boldsymbol{\rho} = \rho_1 Gender_i + \sum_{j=2}^{21} \gamma_j Economic\_Activ_{j,i} + \sum_{h=2}^{7} \varphi_h Region_{h,i} + \sum_{\delta=2}^{8} \delta_j Work\_position_{j,i} \quad (12)$$

In this expression:

$Gender_i$ takes value one if the individual is male and zero if the individual is female;

$Economic\_Activ_{j,i}$ ={1,2,3,…,21} is a dummy variable that takes value of 1 when the individual *i* works in: 1) Agriculture and farm production; 2) Extractive industry; 3) Manufacturing industry; 4) Electricity; 5) Water collection, treatment and distribution; 6) Construction; 7) Wholesale and retail; 8) Transport and storage; 9) Accommodation, catering and similar; 10) information and communication; 11) Financial and insurance; 12) Real estate activities; 13) Consulting, scientific, technical and similar; 14) Administrative and support service; 15) Public administration and defense/compulsory social security; 16) Education; 17) Human health and social support; 18) Artistic, show, sports and recreational activities; 19) Other services; 20) Activities of households employing domestic staff and production activities of households for own use; 21) Activities of international organizations and other extra-territorial institutions;

$Region_{h,i}$ = {1,2,3,4,5,6,7} is a dummy variable that takes value 1 if individual *i* is from: 1*) Norte*; 2) *Centro*; 3) *Alentejo*; 4) *Área metropolitana de Lisboa*; 5) *Algarve*; 6) *Açores*; 7*) Madeira*;

$Work\_position_{j,i}$ = {1,2,3,4,5,6,7,8} is a dummy variable that has value 1 if individual *i* works as (from highest to lowest work position): 1) *Quadros superiores* ; 2) *Quadros médios*; 3)*Encarregados, contramestres, mestres e chefes de equipa*; 4) *Profissionais altamente qualificados*; 5) *Profissionais qualificados*; 6) *Profissionais semiqualificados*; 7) *Profissionais não qualificados*; 8) *Praticantes e aprendizes*. For all dummies, the criteria of excluding one variable in order to avoid problems of collinearity is maintained.

The econometric problems regarding the earnings function are well established in the literature. Because the variance of the logarithm of hourly real wages is not constant across years of schooling, it is expected to observe heteroskedasticity in the variance of error terms using an OLS estimation. To fix this problem, we estimate a regression with robust errors using the package sandwich in R, developed by Zeileis (2004), that estimates a heteroskedasticity consistent (HC) variance-covariance matrix. Another factor emphasized by Campos and Reis (2017) is the omitted variable bias when doing an OLS regression when unobserved characteristics of the individuals, such as ability, are present and are correlated with schooling levels.

From this econometric specification is possible to calculate the relative rates of return to education for each schooling level in the following way: i) those with nine years of schooling relative to those with less than nine years (β_2); ii) those with secondary education relative to those with nine years of schooling ($\beta_3 - \beta_2$); iii) those with post-secondary education relative to those with secondary education ($\beta_4 - \beta_3$); and iv) those with college or more relative to those with secondary education ($\beta_5 - \beta_3$).

Estimating the parameters of the above Mincer Squares (OLS), especially the wage premiums, is one of the main purposes of the current research but is not the only one. On the contrary, another important purpose is to decompose for selected years the variations in the average hourly real earnings as a result of variations in the averages of the explanatory variables and their respective beta estimates. One widely used method of decomposition of the mean values of the dependent variable is the Oaxaca-Blinder decomposition. Following Hlavac (2018), the difference in the average values of the same variable between two distinct groups of data can be expressed as bellow, where $\bar{Y}_A$ is the mean value for group A and $\bar{Y}_B$ is the mean value for group B:

$$\Delta\bar{Y} = \bar{Y}_A - \bar{Y}_B \ (13)$$

From standard econometrics, it is know that the mean value of a variable is also a function of the mean value of a vector of variables and that, by assumption, the mean value of the error term is equal to zero. Therefore, it is possible to express the previous equation as bellow, where $\bar{X}_A^T$ and $\bar{X}_B^T$ are the mean values of the explanatory variables of each group while $\hat{\beta}_A$ and $\hat{\beta}_B$are the associated estimators:

$$\Delta\bar{Y} = \bar{X}_A^T\hat{\beta}_A - \bar{X}_B^T\hat{\beta}_B \ (14)$$

The above mentioned expression can be decomposed, through algebra manipulation, into differences in the mean value of the vector of explanatory variables and differences in the estimators between the two groups, as follows:

$$\Delta\bar{Y} = (\bar{X}_A - \bar{X}_B)^T\hat{\beta}_B + \bar{X}_B^T(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}_A - \bar{X}_B)^T(\hat{\beta}_A - \hat{\beta}_B) \quad (15)$$

In this expression, the mean value of the dependent variable is decomposed in i) differences between the mean values of the vector of explanatory variables for each group; ii) differences in the estimators of each group; and iii) an interaction term between both differences. Using this methodology, it is possible to understand the sources of the variation in the logarithm of hourly real wages and how much each regressor, most importantly schooling levels, and their respective estimates, especially the returns to each level of education, contribute to the changes observed in data.

4. Empirical Analysis

In this section we describe the main attributes of the distribution of Human Capital in Portugal in the last 30 years. We also apply the methodologies described previously to analyze the structural changes in the inequality of Human Capital for Portugal and its participation in income inequality. We also estimate the returns to education for different levels of schooling between 2010 and 2015 and decompose the variation of average hourly real wages in terms of the explanatory variables of the modified Mincerian earnings equation and its coefficients.

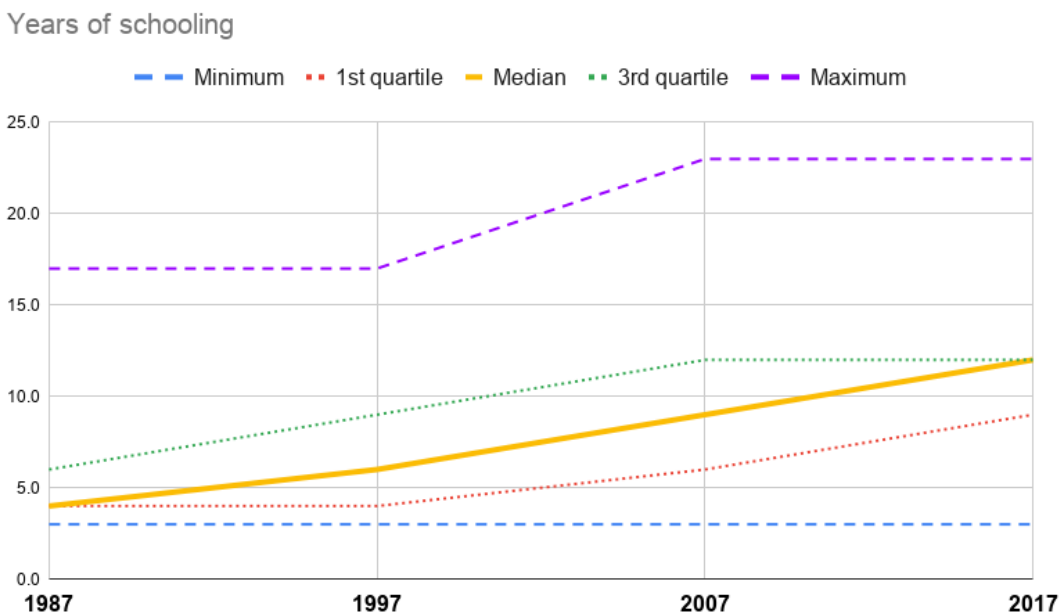
4.1. Distribution of Human Capital in Portugal for the years 1987, 1997, 2007 and 2017

The chosen four points in time to analyze the behavior of Human Capital inequality is not random. First of all, the 10 years in between each of them is important to guarantee that this research captures structural changes and not short-term movements in inequality measurements. Secondly, the last 30 years are one of the most important for Portugal in terms of advancements in education policies that had major impacts over the composition of schooling levels in the private labor market.

In 1987, the distribution of Human Capital throughout the workers in the private sector was heavily concentrated around the 4th year of schooling, with the first and second quartiles of the distribution equalizing that number. Along the subsequent decade, it is already visible a change in the distribution shape, with the second quartile now being the 6th year of schooling and the third quartile going from the 6th to 9th year of schooling.

That change represents a shift to the right in the distribution median, showing improvements in the amount of workers with higher levels of education. In the following decades, the distribution of Human Capital continued to change in a positive direction, with the median of the distribution going from the 6th to the 9th year of schooling and the third quartile increasing from the 9th to the 12th year of schooling while the maximum years of schooling attainment going from 17 to 23 years in 2007. In 2017, the last year for which there is information about the workers' level of education, improvements continue to show up: the second quartile increased from the 9th to the 12th year of schooling although there were no changes in the third quartile, that remained at the 12th year of schooling.

Figure 1. Summary statistics of Human Capital in the Portuguese private labor market over time



Source: Author's own elaboration using *Quadros de Pessoal* dataset.

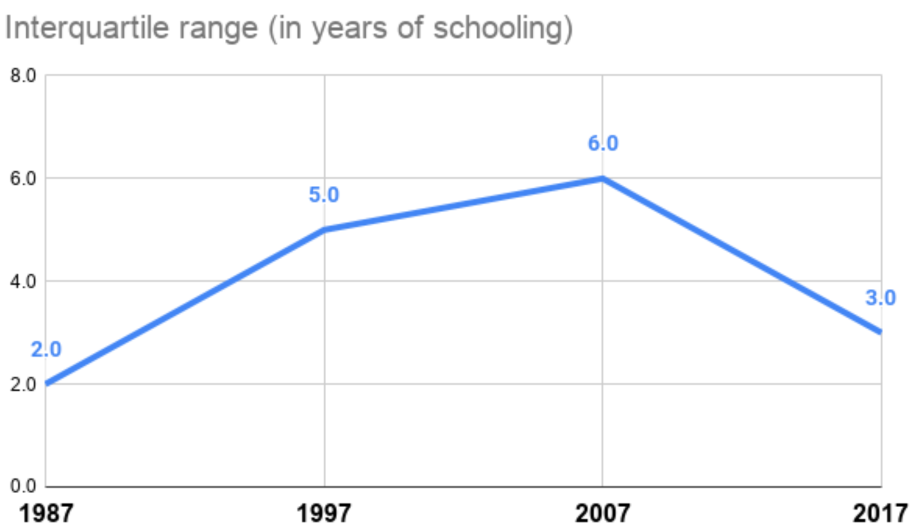
Analyzing the transformation through which the Human Capital distribution went during the last 30 years, it is possible to conclude that the standard worker has had a substantial educational upgrade, going from 4 to 12 years of formal education. The evolution of the first and third quartiles are also positive, with the first one going from 4 to 9 years of schooling and the second one from 6 to 12 years of schooling. In terms of the shape of the distribution, it is possible to say that there was a monotonic shift to the right in the middle of the distribution, meaning that workers pertaining to the subset defined by the interquartile range experienced major changes in their education levels.

The country came out of a situation where the probability mass of the distribution of Human Capital was concentrated around the 4th year of schooling, the left of the distribution, to one where it is concentrated around the 12th year of schooling, the right of the distribution.

4.2. Human Capital Inequality

The easiest and most straightforward way of assessing inequality of a distribution is the interquartile range, i.e. the difference between the third and first quartiles of a given distribution. When this simple statistical procedure is applied to Human Capital, it is possible to observe that it goes from 2 years of schooling in 1987, to 5 years in 1997, reaches a maximum of 6 years in 2007 and goes down to 3 years of schooling by the end of 2017.

Figure 2. Interquartile range of Human Capital distribution in the Portuguese private labor market



Source: Author’s own elaboration using *Quadros de Pessoal* dataset.

Computing the most frequently used inequality indicators for the distribution of Human Capital in the sample of private sector workers, such as the Gini coefficient, the Atkinson index, the Theil T index and the mean error for the logarithmic distribution (Theil N), they all show an ascending trajectory until 2007, while, in the last decade of the analysis, they all show a decrease.

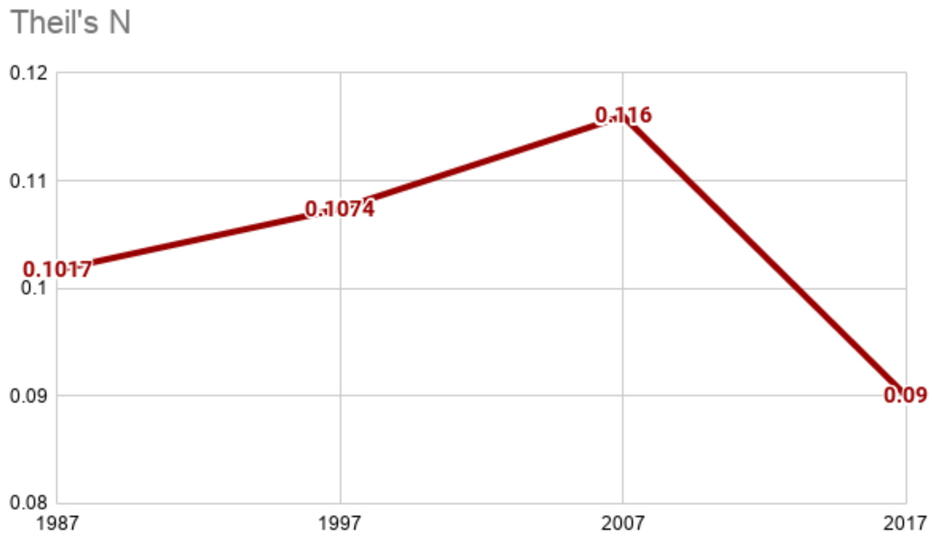
The first fact that stands out is the convergence between different measures of inequality towards a specific trend. This is important because different measures of inequality give different weights to different parts of the distribution. So, when they all converge in that sense, it means that evidence of an increase or decrease in inequality is more robust.

A preliminary assessment of results from inequality measures suggests that the initial increase in the median of schooling years of the population from 1987 onwards first increased the inequality in Human Capital in Portugal, since it was very concentrated, initially, in the 4th year of schooling. The median and the third quartile of the distribution advanced more rapidly in the first years of analysis, enlarging the difference between the middle and the top of the distribution to the bottom. That process kept happening until at least 2007. In 2017, we have almost the opposite picture, with the inequality levels reducing and the years of schooling concentrating in the right part of the distribution, around the 12th year of schooling. In the last 10 years of analysis, there’s a notorious catching up effect, where the first quartile starts to approximate the median, and the median catches up with the third quartile, pushing the inequality downwards regardless of the measure used.

Even though the different measures agree on the shape of the evolution of inequality in time, they differ from each other regarding the intensity of the movements. The measures that attribute relatively higher weights to inequalities in the middle and at the right part of the distribution, such as the Gini, Theil T and the Atkinson ($\varepsilon = 0.5$), tend to show a very small upward movement of inequality until 2007 and a rather intense downward trajectory from 2007 to 2017. This relative smoothness in the upward movement reflect the fact that inequalities increased at a lower speed in the middle and the top part of the distribution. At the same time, indicators that attribute higher weights to the left side of the distribution, such as Theil N and Atkinson with higher values for ε (1 or 2 in the case of this research) capture a more rapid increase in inequality until 2007. Again, these last kinds of indices show that workers with less than the median years of schooling were still relevant in the composition of the work force and were being left behind by the inflow of new and more educated workers[4].

[4] For all Human Capital inequality indices calculated see figures A.3-A.6 in the Appendix.

Figure 3. Inequality of Human Capital in the Portuguese private labor market measured by the Theil's N index



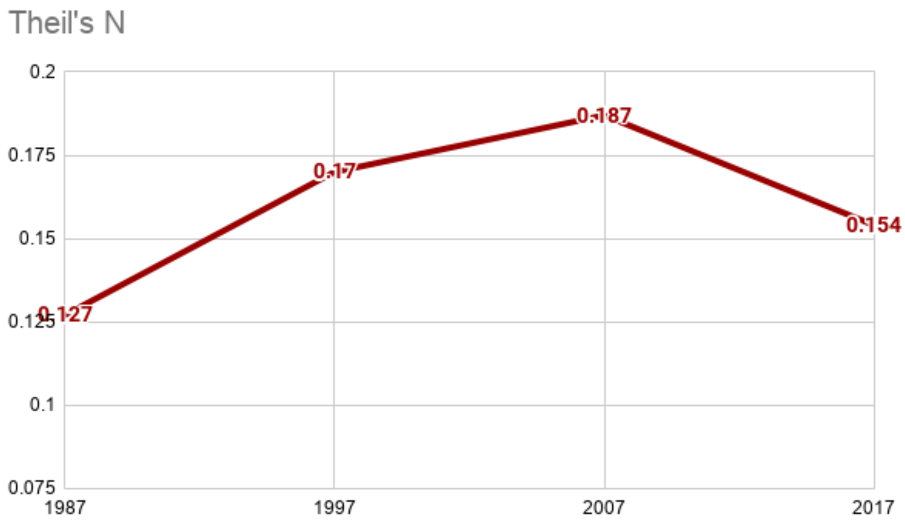
Source: Author's own elaboration using *Quadros de Pessoal* dataset.

As a by-product of this analysis, we see that the dynamics of the main statistical moments of the distribution suggests a Kuznets curve for Portuguese Human Capital, a relationship that was already verified in previous studies of inequality of Human Capital in the country, such as in Duarte, Fidalgo & Simões (2010) mentioned in section 2.3., in which inequality first increases along with the average schooling years and, after reaching a maximum, starts falling as the schooling years keep on growing, forming an inverted parabola. This behavior tends to be associated to the age composition of the stock of workers in the private sector. Due to the mandatory increase in years of education, starting in 1964 with 6 years of formal schooling, going to 9 years in 1986 and reaching 12 years in 2009 according to the Portuguese Ministry of Education, it is reasonable to infer that the inflows of individuals with higher levels of schooling started driving inequalities in Human Capital upwards to the point where the distribution of Human Capital, with older people leaving the labor market, became more homogenous, with less differences in intergenerational years of schooling.

4.3. Decomposition of earnings inequality by differences of educational attainment

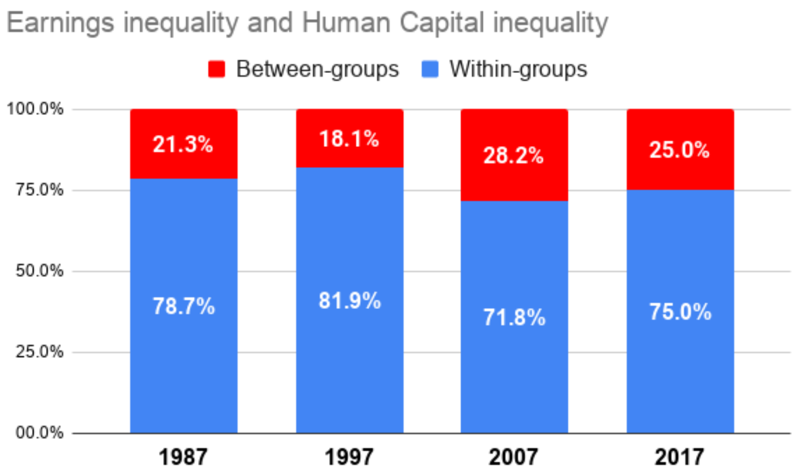
In terms of decomposition of inequality indices, it is only possible to additively decompose indexes that are of the class of Generalized Entropy Indexes (GEI), such as Theil's N, also known as the Logarithmic Mean Error. This decomposition tells us how much of the static inequality of earnings comes from within and between-group inequality, i.e. what's the contribution of differences in income in a certain group (e.g. the variability in income for workers that hold secondary education), and the contributions of differences between the mean income among the different groups that constitute the population to the total earnings inequality. The temporal path of earnings inequalities, calculated using Theil's N methodology, is shown in figure 4. The idea is to know, for each of those levels of inequality, how much of it comes from Human Capital inequalities (between-group) and how much comes from other factors (within-group)

Figure 4. Inequality of earnings in the Portuguese private labor market measured by the Theil's N index



Source: Author's own elaboration using *Quadros de Pessoal* dataset.

Figure 5. Additive decomposition of the Theil's N index for inequality of income conditioned on variability in Human Capital



Source: Author's own elaboration using *Quadros de Pessoal* dataset.

The variable used for the purpose of group identification was the years of schooling of the workers, as an approximation for Human Capital groups. For all the years at stake, the majority of the static inequality comes from within group inequality, ranging from 78.7% to 81.9%, showing evidence that factors beyond educational attainment have an important influence over earnings inequality as noted by Rodrigues (1996). At the same time, between 18.1% and 28.2% of earnings inequality comes from between group inequalities and shows that differences in educational attainment are an important factor associated with differences in earnings. Rodrigues (1996) also showed in his decompositions that the major variable associated with earnings inequality is the disparities in years of schooling when compared to regional or even gender inequalities.

But one of the main purposes of the investigation is to build and analyze the evolution and behavior of these contributions over time in order to better explain the earnings inequality profile. From 1987 to 1997, approximately 92.5% of the increase in earnings inequality came from changes in the within group component, while the other 7.5% came from changes in between group inequality. The total participation of Human Capital inequality in earnings inequality fell from 21.3% to 18.1%. From 1997 to 2007, the roles were inverted, with the between group component accounting for the vast majority of the increase in earnings inequality level. The within group component even registered a negative contribution, even if only slightly. In that decade, the participation of Human Capital inequality in earnings inequality rose to 28.2%. From 2007 to 2017, the decreasing inequality was attributed in a more balanced way to within and between group inequalities. The former explaining almost 57.0% of the decrease, while the latter explaining 43.0%.

With the information above, it is not possible to determine a trend for between-group inequalities due to its fluctuation over the decades. Apart from that, it is possible to make two statements: one can say that differences in Human Capital were relatively more associated to earnings inequality in the last two decades of analysis and that Human Capital inequality participation on earnings inequality reduced in the last decade of analysis. To understand better the dynamics of this relationship, it is also important to explore another important dimension affecting the link between Human Capital inequalities and earnings inequalities: the returns to education.

4.4. Estimation of returns to levels of schooling in R for 2010 and 2015 using a modified Mincerian based earnings function

A series of Mincerian regressions were estimated using data for the Portuguese private labor market and also for the regional private labor market, using the well established NUTS2 criteria. Our econometric analysis uses 2010 and 2015 data points, since this is the most spaced period for which there is all the information needed to comply with the Mincerian earnings specification and all the controls added to the model. But data constraints above mentioned do not mean the selected years do not fulfill suitable macroeconomic criteria, on the contrary. 2010 and 2015 are comparable in macroeconomic terms since the two capture the Portuguese economy growing at roughly the same pace, which serves as an additional control variable for the economic cycle.

The first regression was estimated for the whole Portuguese private labor market for the years 2010 and 2015, using the OLS estimator with robust errors since the null hypothesis of the Breusch-Pagan test (the error term presents homoscedasticity) was rejected for both years at the 1.0% level. This means that it was necessary to adjust the conventional estimation method to incorporate robust errors. The specification of the equation took into consideration all the control variables already mentioned:

$$\boldsymbol{Z^T}_{\boldsymbol{i}}\boldsymbol{\rho} = \rho_1 Gender_i + \sum_{j=2}^{21} \gamma_j Economic\_Activ_{j,i} + \sum_{h=2}^{7} \varphi_h Region_{h,i} + \sum_{\delta=2}^{8} \delta_j Work\_position_{j,i} \quad (16)$$

Table 1. shows the more important part of the regression output for both years. First, all the estimators have the signs expected by the theoretical literature, according to section 2.1. All the main estimators are positive, with the exception of the quadratic term associated with experience, confirming diminishing returns to additional years of on the job-training. Secondly, the returns to each level of schooling follow the same behavior as the ones found in Campos and Reis (2017), with returns to higher education decreasing in recent years. The estimations undertaken in this research show that the private wage premium for workers that completed nine years of education in 2010 is 12.9% with respect to those that have less than nine years of education, while the wage premium associated with secondary schooling is 13.4% compared to those with nine years of schooling and the one associated to a college degree or more is 31.6% in relation to those with secondary schooling. Campos and Reis (2017) obtain returns

of more than 45.0% for those with tertiary education, although with the same decreasing characteristics as in our regressions. One possible reason why there is such differences between the estimators of returns to schooling according to both researches is that our regression controls levels of education for more factors, such as the type of economic activity, region and especially the work position, which tends to mitigate education weight on disparities of income in cross-sectional data.

Removing the control for the work position, for instance, increases the returns to schooling of individuals with a college degree or more with respect to those with secondary education in roughly 25 percentage points, from 31.6% to 56.8%[5], which overestimates the parameters associated to the real wage premium of higher education. For 2010, the number of workers with secondary education that occupied high posts in their company is not at all trivial. For instance, considering four of the highest work positions available in the QP data base, namely: i) *Quadros Superiores*; ii) *Quadros Médios*; iii) *Encarregados, Constramestres, Mestres e Chefe de Equipa* and iv) *Profissionais Altamente Qualificados*, the participation of individuals with secondary education in these categories is respectively: 37.6%, 25.7%, 23.9% and 13.3% for 2010.

That means that there are still a lot of individuals with relatively low skill in well paid high work positions. One possibility is that they are capital owners. If it is intended to understand the relationship of inequalities in Human Capital and inequalities in real earnings, this factor seems to be an important one to be addressed when calculating these returns. Beyond that, the dynamics of these returns can be an important factor behind the falling weight of inequalities in Human Capital in differences in real income between individuals, as pointed out previously when we computed the decomposition of Theil's N index of earnings inequality.

For 2015, it is possible to observe that the returns fall across all levels of education, most drastically for those with nine years of education (-17.2%), for those with post-secondary education (-16.2%) and those with a college degree or more (-13.9%). The individuals with secondary education were the ones with the smallest loss (-8.2%). This result agrees with the income inequality decomposition done in section 4.3., where, in the last decade, it is possible to observe that differences in schooling levels have been less associated with differences in income. It is true that inequalities in years of schooling have decreased between 2007 and 2017, and that for itself helps to explain the smaller influence of differences in Human Capital on earnings inequality, but the reduced relative returns to higher education also contribute to speed

[5] Regression table A3. in the Appendix

up this process. Another evidence of a convergence in real wages across schooling levels can be obtained by comparing the change in average hourly real wages between 2010 and 2015. For those with 12 years of formal education, hourly real wages drop 11.2% while for those with a college degree or more this number becomes more negative, reaching 14.1%.

Although the analysis between only two points in time may take away information about how these returns have been fluctuating in the period, the negative trend of returns to higher education in Portugal has been noticed by Almeida *et al* (2017). One of the reasons pointed out by their research is that individuals with a college education with skills or characteristics that are not valued by the market, or even perform tasks that don't use all of their knowledge, may have a lower and decreasing wage premium.

Table 1. Earnings regression controlled by gender, region, work position and economic activity using OLS with robust errors

	(2010)	(2015)
(Intercept)	1.09*** (1.0373e-01)	1.06*** (3.6890e-02)
Exp	1.93e-02*** (7.79114e-05)	1.79e-02*** (7.4439e-05)
Exp ²	-2,9e-04*** (1.3854e-06)	-2.54e-04*** (1.3266e-06)
Tenure	1.11e-02*** (3.4329e-05)	1.15e-02*** (3.2255e-05)
Ninth_grade	1.29e-01*** (6.0856e-04)	1.06e-01*** (6.2432e-04)
Secondary	2.62e-01*** (7.7223e-04)	2.29e-01*** (7.5746e-04)
Post_secondary	4.66e-01*** (1.9939e-03)	4.00e-01*** (1.9519e-03)
College or more	5.78e-01*** (1.3231e-03)	5.01e-01*** (1.2086e-03)
Control variables(…)[6]		
Residual S.E.	0.3301	0.3220
Adjusted R-squared	0.6099	0.6013
P-value(F)	< 2.2e-16	< 2.2e-16

Source: Authors' own calculations using R software and *Quadros de Pessoal* dataset. Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ² is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \dots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model

[6] For the complete regression table see Table A1. In the Appendix

specified does better than the intercept to account for the regressand variability.

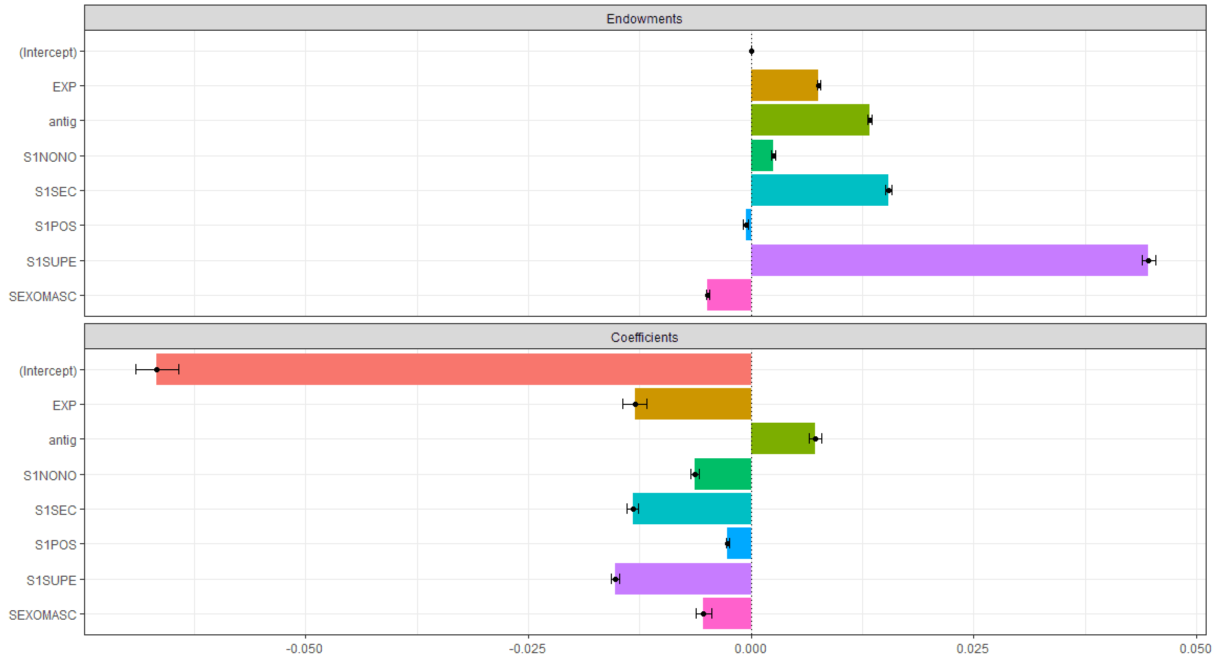
In what follows we present the results from the threefold Oaxaca-Blinder decomposition of the log of real hourly earnings using the modified Mincerian earnings regression function with OLS estimators. The control variables used were restricted to tenure and gender, which is commonly used in the earnings function economic literature.

Table 2. Threefold Oaxaca-Blinder decomposition of the log of real hourly earnings using the Mincerian earnings regression function with OLS estimators

	(Endowments)	**(Coefficients)**
(Intercept)	-	-0.0666*** (1.239822e-03)
Exp	0.00758*** (8.901533e-05)	-0.0130*** (6.853500e-04)
Tenure	0.01334*** (1.277517e-04)	0.0072*** (3.501025e-04)
Ninth_grade	0.00253*** (1.158450e-04)	-0.0063*** (2.435064e-04)
Secondary	0.01545*** (1.858477e-04)	-0.0133*** (3.091488e-04)
Post_secondary	-0.00061*** (1.392081e-04)	-0.0026*** (7.033684e-05)
College or more.	0.04462*** (3.676523e-04)	-0.0152*** (2.402758e-04)
Control variables (...)		

Source: Author'sown calculations using R software (package oaxaca) and *Quadros de Pessoal* dataset. Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. The values under "Endowments" represent the contribution of the variation of each regressor to the total variation of the dependent variable, while the values under "Coefficients" represent the contribution of the variation in each coefficient to the total variation of the dependent variable. Because the dependent variable is measured in log, the contributions are all in percentage points when multiplied by 100.

Figure 6. Threefold Oaxaca-Blinder decomposition: how much endowments and coefficients contributed to the decline in average hourly real wages between 2010 and 2015



Source: Authors' own calculations. See table 2. for a detailed description of the Oaxaca-Blinder components.

Between 2010 and 2015, the mean hourly real wages fell by 4.3%. The change in the explanatory variables during 2010 and 2015 are all positive across experience, tenure and the composition of education among workers. Ceteris paribus, this positive variation in endowments would contribute to an increase of 7.9 percentage points in hourly real wages. It is important to note that from the 7.9 percentage points of contribution given by variations in the endowments to the average hourly real wages, 4.6 percentage points, or 58.2% of that total amount, come only from positive variations in the amount of workers with a college degree or more. The increased number of workers with secondary education also led to a significant positive pressure over average hourly real wages of 1.5 percentage points. Given that changes in the dependent variable through time are not only driven by changes in endowments, it is necessary to account for the changes in the coefficients. The total, ceteris paribus, change in average hourly real wages from variations in the coefficients is around -11.5 percentage points and more than offset the positive variations brought about by the endowments. It is noteworthy, nonetheless, to address the fact that a large part of this negative amount, almost 60%, comes from variations in the intercept of the underlying earnings function. The explanation for this may be related to exogenous factors affecting the earnings function, such as the decrease in the real minimum wage in the period, of almost 1.0%, and other institutional factors related to the stabilization program Portugal was subjected after the 2011 European debt crisis. The institutional transformations with short-term negative effects on Gross Domestic Product (GDP) through which Portugal has gone during 2010 and 2015 is an important candidate to explain why wage premiums decline across all levels of formal schooling. As a separate category, it is also interesting to notice the fall in the relative participation of men in the work space, a negative driving force of average hourly real wages. Combined with lower gender wage premium, this is a positive factor for the gender equality perspective in the Portuguese private labor market.

When comparing the contributions of endowments and coefficients to the variations of average hourly real wages, it is possible to conjecture that public policies successfully aimed at reducing Human Capital inequalities in the last decade may be associated with declining wage premiums through a supply side effect in the labor market. With technological driven demand for labor increasing at a lower speed than the supply of higher skilled workers, the real wage equilibrium is expected to fall as a result, as shown in Centeno and Novo (2014).

Apart from the exogenous contribution to the decrease in hourly real wages, the decomposition shows clearly that the decrease in the returns associated with a college degree or more was the main identifiable force compressing average hourly real wages, with a contribution of 31.4% of the 40.0% left to be explained. The other part comes mainly from falling returns to secondary education and experience, amounting altogether to roughly 90.0% of the identifiable 40.0%. It is also worth mentioning that this broad picture of declining wage premiums for higher education in Portugal during 2010 and 2015 is a reflection of regional dynamics, as can be seen in tables A6. and A6. of the appendix. The average return on higher than secondary education fell for all seven regions analyzed. Alentejo and Açores registered the highest decrease in those returns, respectively, -25.0% and -23.0%, while Lisboa faced the lowest one, around -8.0%.

5. Concluding remarks

In this research, we characterize Human Capital distribution in Portugal and try to assess its importance in the earnings distribution for Portugal. It was shown that the average Portuguese in the private labor market has seen his educational attainment increase monotonically from 4 to 12 years of schooling. As was already noted, this work takes the measure of years of schooling as proxy to Human Capital, without any adjustments to the quality of the learning process in formal education. Nonetheless, this increased school attainment has been accompanied by an increase in the inequality of Human Capital, specifically between 1987 and 2007. The reason for this is the change in the composition of the private labor market, with newcomers with more years of schooling, on average, than the ones before. That is an achievement of the mandatory completion of increasing years of schooling established by the Portuguese government, beginning with 6 years in the 1960’s, 9 years in the 1980’s and 12 years in the 2000’s. By the time the effects of these impositions reduced the difference in education between generations, the inequality of Human Capital started to respond positively, falling between 2007 and 2017. This behavior altogether presents an inverted parabola shape to the inequality of Human Capital in the Portuguese private labor market during a period of 30 years and is an important achievement, since this type of inequality is negatively associated to economic growth when financial markets fail to allocate resources optimally.

When analyzing how significant the Human Capital inequality pattern has been to inequality of before tax earnings, it is clear that it is an important issue. The contribution of Human Capital inequality to inequality of income fluctuates around 18.1% and 28.2% in the period analyzed, which, for a single factor, is a lot. As seen in section 4.3., there is no clear trend for this indicator. Therefore, particularly for this research, we are interested in analyzing what are the possible reasons associated with the decreasing participation of Human Capital inequality in earnings inequality in the last decade.

One possible factor why the weight of Human Capital inequality on earnings inequality reduced between 2007 and 2017 is due to the process of convergence in years of schooling among workers in the private sector, meaning that these shorter educational distances were accompanied by higher average earnings, which has been a positive force driving the reduction of earnings inequalities.

But analyzing the behavior of inequalities in Human Capital by itself is only one part of the story, since the economic returns to additional years of education can also be a force driving the distribution of earnings among workers. And that is what was done in section 4.3. Even with the mentioned restrictions of the dataset to perform the regressions necessary to estimate the rate of private returns to education for all the sample, the years selected to conduct, i.e. 2010 and 2015, the modified Mincerian estimations can give an idea of how these returns behaved over time. It is clear from the estimations conducted in section 4.3. that the return to higher education, meaning post-secondary and college or more, was the one that fell most when considering the relevant part of the distribution of Human Capital. That means that the wage premium of higher education relative to secondary education decreased during 2010 and 2015, respectively 16.2% and 13.9%, and that rapidly compressed wage premiums in that level of education may explain why inequalities in Human Capital can have a lower participation in the inequality of earnings in the recent decade.

If that is the case, reduced higher education wage premiums as a driving force for the reductions in earnings inequalities does not seem to be - differently than reduction in the inequalities of Human Capital - a good way to shrink the disparities in earnings. As documented in the literature, these reduced wage premiums may come from a mismatch in the labor market between abilities of the worker and the needs of the employer, like in Campos and Reis (2017). In this regard, further investigation can be conducted to understand if the mismatch is coming from the supply side, with students with higher education not having a good quality of learning, or from the demand side, with companies not having the technological capabilities to take advantage of the skills of these workers. For further research, the current work can provide ways of adjusting years of schooling to returns to education in order to create more interesting measures of Human Capital for the Portuguese economy.

6. References

Aghion, P., Ljungqvist, L., Howitt, P., Howitt, P. W., Brant-Collett, M., & García-Peñalosa, C. (1998). *Endogenous growth theory*. MIT press.

Arnold, J., & Farinha Rodrigues C. (2015). *Reducing Inequality and Poverty in Portugal* (OECD Economics Department Working Papers, No. 1258).

Barro, R. J., & Lee, J. W. (1993). *International comparisons of educational attainment* (National Bureau of Economic Research Working Papers No. 4349).

Barro, R., & Lee, J. (2001). Schooling Quality in a Cross Section of Countries. *Economica*, *68*(272), 465–488.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New-York: University of Chicago Press.

Benabou, R. (1996). Inequality and growth. In Ben S. Bernanke & Julio J. Rotemberg (Eds.), *NBER Macroeconomics Annual*, 11, (pp.11-74). Massachusetts: MIT Press.

Campos, M., & Reis, H. (2017). Revisiting the returns to schooling in the Portuguese economy. *Banco de Portugal Economic Studies,* 3(2), 1-28.

Centeno, M., & Novo, Á. A. (2014). When supply meets demand: Wage inequality in Portugal. *IZA Journal of European Labor Studies*, 3(23), 1-20.

Chen, M., Chernozhukov, V., Fernández-Val, I., & Melly, B. (2017). Counterfactual: An R Package for Counterfactual Analysis. *The R Journal*, *9*(1), 370. https://doi.org/10.32614/rj-2017-033

Cowell, F.A. (2000). Measurement of inequality. In Anthony B. Atkinson & François. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 1, chapter 2 (pages 87-166). London: Elsevier.

Cowell, F. (2011). *Measuring Inequality*, 3rd edition. London: Oxford University Press.

Emamian, M. H. (2019). Blinder-oaxaca decomposition. *Archives of Iranian Medicine*, *22*(3), 164–165.

Fidalgo, J. G., Simões, M., & Duarte, A. (2014). Mind the Gap: education inequality at the regional level in Portugal, 1986-2005. *Notas Económicas*, (32), 22–43. https://doi.org/10.14195/2183-203x_32_2

Figueiredo, H., Portela, M., Sá, C., Cerejeira, J., Almeida, A., & Lourenço, D. (2017). *Benefícios do Ensino Superior*. Lisboa: Fundação Francisco Manuel dos Santos.

Hlavac, M. (2014). oaxaca: Blinder-Oaxaca decomposition in R. Available at SSRN 2528391.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.

Lemieux, T. (2006). The "Mincer equation" thirty years after schooling, experience, and earnings. In Shoshana Grossbard (Ed.), *Jacob Mincer a Pioneer of Modern Labor Economics*, (pp. 127-145). Boston, MA: Springer.

Lucas, R. (1988). On the mechanics of economic development." *Journal of Monetary Economics*, 22 (1), 3–42. https://doi.org/10.1016/0304-3932(88)90168-7

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, *12*(4), 577–590. https://doi.org/10.1016/j.labeco.2005.05.006

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4), 281-302.

Mincer, J. (1974). Progress in Human Capital Analysis of the distribution of earnings. National Bureau of Economic Research.

Mookherjee, D., & Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal*, 92(368), 886-902.

Mulligan, C. B., & Sala-i-Martin, X. (1994). *Measures of the Aggregate Value of Human Capital*. Mimeo, University of Chicago.

Mulligan, C. B., & Sala-i-Martin, X. (1995). A labor-income-based measure of the value of human capital: an application to the states of the United States (National Bureau of Economic Research, Working Papers No.5018).

Pereira, J. P. C. (2005). Measuring human capital in Portugal. *Notas Económicas*, 21, 16-34.

Polachek, S. W. (2008). Earnings over the life cycle: The Mincer earnings function and its applications. *Foundations and Trends in Microeconomics*, *4*(3), 165-272.

Rodrigues, C. F. (1996). Medição e Decomposição da Desigualdade em Portugal (1980 - 1990). *Revista de Estatística*, *3*(1), 47–70.

Rodrigues, C. F. (2008). *Distribuição do Rendimento, Desigualdade Pobreza: Portugal nos anos 90*, Coimbra: Edições Almedina.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2), 407-437.

Savvides, A., & Stengos, T. (2008). *Human Capital and Economic Growth*, Standford: Stanford University Press.

Schultz, T. W. (1960). Capital formation by education. *Journal of Political Economy*, *68*(6), 571-583.

Sørensen, P. B., & Whitta-Jacobsen, H. J. (2005). *Introducing advanced macroeconomics: growth and business cycles*. London: McGraw-Hill Education.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators (*Research Report Series / Department of Statistics and Mathematics*, Nr. 10).

Appendix

Table A1. Conversion of level of schooling to years of schooling

Level of schooling	Years of schooling (before Bologna)	Years of schooling (after-Bologna)
Inferior ao 1∘ ciclo do Ensino Básico	3	3
1∘ ciclo do Ensino Básico	4	4
2 ∘ *ciclo do Ensino Básico*	6	6
3 ∘ *ciclo do Ensino Básico*	9	9
Ensino Secundário	12	12
Ensino pós-secundário	-	13
Bacharelato	15	14
Licenciatura	17	17
Mestrado	-	19
Doutoramento	-	23

Source: Author's conversion table based on data from the European Comission and DGES. Notes: - means that there were no observations in the dataset with those characteristics.

Table A2. Earnings regression controlled by gender, region, work position and economic activity using OLS with robust errors

	(2010)	**(2015)**
(Intercept)	1.09*** (1.0373e-01)	1.06*** (3.6890e-02)
Exp	1.93e-02*** (7.79114e-05)	1.79-02*** (7.4439e-05)
Exp ^{2}	-2.79e-04*** (1.3854e-06)	-2.54e-04*** (1.3266e-06)
Tenure	1.11e-02*** (3.4329e-05)	1.15e-02*** (3.2255e-05)
Ninth_grade	1.29e-01***	1.06e-01***

	(6.0856e-04)	(6.2432e-04)
Secondary	2.62e-01*** (7.7223e-04)	2.29e-01*** (7.5746e-04)
Post_secondary	4.66e-01*** (1.9939e-03)	4.00e-01*** (1.9519e-03)
College or more	5.78e-01*** (1.3231e-03)	5.01e-01*** (1.2086e-03)
Dummy_Alentejo	-1.5978e-03*** (1.7827e-03)	-7.8390e-03*** (1.8464e-03)
Dummy_Algarve	2.5741e-02*** (1.9090e-03)	1.4467e-02*** (1.9453e-03)
Dummy_Centro	-4.0786e-02*** (1.5744e-03)	-5.5743e-02*** (1.6515e-03)
Dummy_Lisboa	1.1475e-01*** (1.5563e-03)	8.5177e-02*** (1.6325e-03)
Dummy_Madeira	7.4195e-02*** (2.0212e-03)	5.5050e-02*** (2.2209e-03)
Dummy_Norte	-5.4880e-02*** (1.5409e-03)	-5.7012e-02*** (1.6204e-03)
Dummy_Aprendiz	-3.1580e-01*** (1.2813e-03)	-2.8753e-01*** (1.3427e-03)
Dummy_Chefe_Equipa	9.9906e-02*** (1.4385e-03)	1.2335e-01*** (1.4585e-03)
Dummy_Quadro_Medio	1.3372e-01*** (1.4871e-03)	1.5339e-01*** (1.4600e-03)
Dummy_Nao_Qual	-3.0886e-01*** (1.1304e-03)	-2.9995e-01*** (1.1042e-03)
Dummy_Qualificado	-2.1504e-01*** (9.8721e-04)	-2.0248e-01*** (9.7636e-04)

Dummy_Semi_Qual	-2.9757e-01*** (1.0429e-03)	-2.7999e-01*** (1.0236e-03)
Dummy_Quadro_Superior	2.7128e-01*** (1.6115e-03)	2.9223e-01*** (1.5618e-03)
Dummy_Homem	1.5347e-01*** (4.9952e-04)	1.4563e-01*** (4.9031e-04)
Dummy_ Extractive industry	1.8483e-01*** (4.1741e-03)	2.4934e-01*** (4.9965e-03)
Dummy_ Manufacturing	6.8898e-02*** (1.8688e-03)	8.7816e-02*** (1.7309e-03)
Dummy_ Electricity	5.4752e-01*** (3.5355e-03)	5.9598e-01*** (3.6681e-03)
Dummy_ Water collection	1.5384e-01*** (2.8764e-03)	9.8533e-02*** (2.6003e-03)
Dummy_ Construction	1.3133e-02*** (1.9482e-03)	1.5168e-02*** (1.9041e-03)
Dummy_ Wholesale and retail	3.6751e-02*** (1.8878e-03)	3.4719e-02*** (1.7546e-03)
Dummy_ Transport and storage	1.8725e-01*** (2.1337e-03)	1.6406e-01*** (2.0124e-03)
Dummy_ Accommodation	-9.4163e-02*** (1.9693e-03)	-8.1891e-02*** (1.8221e-03)
Dummy_ Financial and insurance	1.8317e-01*** (2.5390e-03)	1.3253e-01*** (2.3471e-03)
Dummy_ Real estate activities	5.1404e-01*** (2.2689e-03)	4.8692e-01*** (2.1923e-03)
Dummy_ Consulting, scientific, technical and similar	-3.3976e-02*** (3.8688e-03)	-4.3961e-02*** (3.6706e-03)

Dummy_ Administrative and support service insurance	7.5055e-03** (2.2980e-03)	-5.7325e-03** (2.1449e-03)
Dummy_ Public administration and defense/compulsory social security	5.7589e-02*** (2.0353e-03)	5.1231e-03** (1.8916e-03)
Dummy_ Education	9.0185e-02*** (3.2930e-03)	5.2882e-02*** (3.2541e-03)
Dummy_ Human health and social support	1.0196e-01*** (2.6227e-03)	3.9666e-02*** (2.4330e-03)
Dummy_ Artistic, show, sportsand recreational activities	1.9725e-02*** (1.9290e-03)	4.8702e-03** (1.7752e-03)
Dummy_ Other services	1.1840e-01*** (3.5840e-03)	8.1320e-02*** (3.4675e-03)
Dummy_ Activities of households employing domestic staff	-6.4258e-03** (2.2220e-03)	-2.8542e-03 (2.1591e-03)
Dummy_ Administrative and support service insurance	4.4117e-01*** (3.6604e-02)	4.5452e-01*** (3.2663e-02)

Residual S.E.	0.3301	0.3220
Adjusted R-squared	0.6099	0.6013
P-value(F)	< 2.2e-16	< 2.2e-16

Source: Author's own calculations using R software and *Quadros de Pessoal* dataset. Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ^{2} is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \ldots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model specified does better than the intercept to account for the regressand variability.

Table A3. Earnings regression controlled by gender, region and economic activity (not by work position) using OLS with robust errors

	(2010)	**(2015)**
(Intercept)	0.68*** (2.7262e-03)	0.66*** (2.6405e-03)
Exp	2.61e-02*** (8.3233e-05)	2.40-02*** (8.0015e-05)
Exp ^{2}	-3.63e-04*** (1.4811e-06)	-3.31e-04*** (1.4238e-06)
Tenure	1.37e-02*** (3.700e-05)	1.44-02*** (3.4582e-05)
Ninth_grade	1.75e-01*** (6.5353e-04)	1.44e-01*** (6.661e-04)
Secondary	3.76e-01*** (8.2962e-04)	3.28e-01*** (8.185e-04)
Post_secondary	7.51e-01*** (2.133e-03)	6.51e-01*** (2.1496e-03)
College or more	9.44e-01*** (1.1986e-03)	8.44e-01*** (1.1400e-03)
Control variables(…)		

Residual S.E.	0.3583	0.3518
Adjusted R-squared	0.5402	0.524
P-value(F)	< 2.2e-16	< 2.2e-16

Source: Author'sown calculations using R software and *Quadros de Pessoal* dataset. Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ^{2} is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \ldots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model specified does better than the intercept to account for the regressand variability.

Table A4. Mincerian regression controlling for economic activity, work position and gender using regional data for 2010 using OLS with robust errors

	(NORTE)	(ALGARVE)	(CENTRO)	(ALENTEJO)
(Intercept)	1,08e*** (5,297e-03)	1,27e*** (1,1526e-02)	1,09e*** (4,8297e-03)	1,21e*** (8,1397e-03)
Exp	1,74e-02*** (1,3321e-04)	1,20e-02*** (3,7155e-04)	1,54e-02*** (1,7081e-04)	1,36e-03*** (3,233e-04)
Exp ²	-2,34e-04*** (2,3868e-06)	-1,85e-04*** (6,429e-06)	-2,25e-04*** (2,9497e-06)	-2,01e-04*** (5.4212e-06)
Tenure	7,53-03*** (5,3475e-05)	9,20-03*** (1,8871e-04)	8,02e-03*** (7,5279e-05)	1,05e-02*** (1,5852e-04)
Ninth_grade	1,39e-01*** (9,9187e-04)	5,06e-02*** (2,9084e-03)	9,07e-02*** (1,2927e-03)	7,07e-02*** (2,4593e-03)
Secondary	2,77e-01*** (1,3370e-03)	1,49e-01*** (3,8289e-03)	1,90e-01*** (1,6906e-03)	1,73e-01*** (3,2802e-03)
Post.Sec and college	5,85e-01*** (12,2769e-03)	4,28e-01*** (6,8072e-03)	5,31e-01*** (2,9187e-03)	5,27e-01*** (6,2494e-03)

Residuals S.E.	0.3097	0.3106	0.3038	0.3065
AdjustedR-squared	0.5861	0.4268	0.4987	0.5127
P-value(F)	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

Source: Author'sown calculations using R software and *Quadros de Pessoal* dataset.

Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ² is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \dots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model specified does better than the intercept to account for the regressand variability.

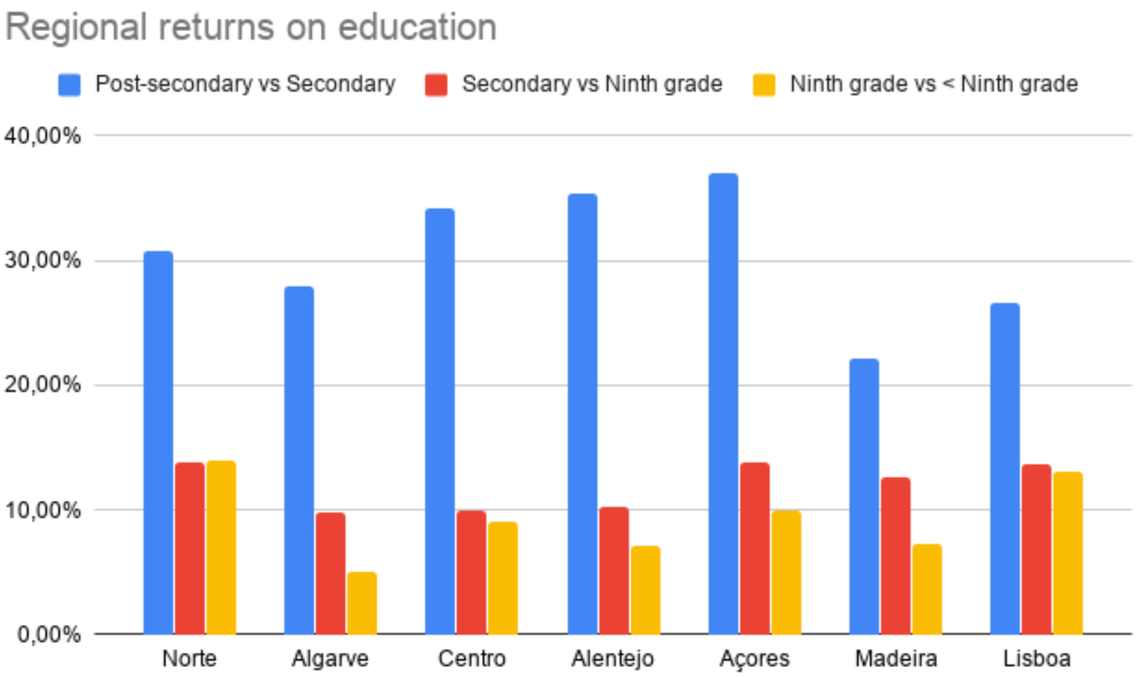
Table A5. Mincerian regression controlling for economic activity, work position and gender using regional data for 2010 using OLS with robust errors

	(AÇORES)	(MADEIRA)	(LISBOA)
(Intercept)	1,19e*** (1,6809e-02)	1,10e*** (1,81586e-02)	1,09e*** (5,6800e-03)
Exp	1,82e-02*** (5,6380e-04)	1,86e-02*** (4,9074e-04)	2,17e-02*** (1,3384e-04)
Exp ^{2}	-2,61e-04*** (1,0271e-05)	-2,80e-04*** (8,6976e-06)	-3,34e-04*** (2,4607e-06)
Tenure	1,13-02*** (2,5555e-04)	1,02e-02*** (2,2451e-04)	1,55e-02*** (6,2536e-05)
Ninth_grade	1,00e-01*** (93,8681e-03)	7,21e-02*** (3,46835e-03)	1,30e-01*** (1,1741e-03)
Secondary	2,38e-01*** (5,5132e-03)	1,98e-01*** (4,6835e-03)	2,67e-01*** (1,3674e-03)
Post.Sec and college	6,08e-01*** (1,1300e-02)	4,20e-01*** (9,3164e-03)	5,33e-01*** (2,0093e-03)

Residuals S.E.	0.2745	0.2876	0.3546
AdjustedR-squared	0.6414	0.6684	0.6287
P-value(F)	< 2.2e-16	< 2.2e-16	< 2.2e-16

Source: Author'sown calculations using R software and *Quadros de Pessoal* dataset. Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ^{2} is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \dots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model specified does better than the intercept to account for the regressand variability.

Figure A1. Regional wage premiums for 2010



Source: elaborated by the author using *Quadros de Pessoal* dataset.

Table A6. Mincerian regression controlling for economic activity, work position and gender using regional data for 2015 using OLS with robust errors

	(NORTE)	(ALGARVE)	(CENTRO)	(ALENTEJO)
(Intercept)	1,04*** (2,2935e-03)	1,27*** (1,1526e-02)	1,02*** (4,4522e-03)	1,15*** (7,5678e-03)
Exp	1,77e-02*** (7,5200e-05)	9,82e-02*** (3,3040e-04)	1,54e-02*** (1,6181e-04)	1,20e-03*** (3,157e-04)
Exp ^{2}	-2,50e-04*** (1,3419e-06)	-1,55e-04*** (5,897e-06)	-2,22e-04*** (2,8185e-06)	-1,71e-04*** (5,4105e-06)
Tenure	1,19-03*** (3,2840e-05)	1,03-03*** (1,7001e-04)	8,55e-03*** (6,9848e-05)	1,13e-02*** (1,5011e-04)
Ninth_grade	1,16e-01*** (6,3036e-04)	3,31e-02*** (2,8405e-03)	9,07e-02*** (1,2927e-03)	6,25e-02*** (2,5552e-03)
Secondary	2,46e-01*** (7,6385e-04)	1,23e-01*** (3,4796e-03)	1,68e-02*** (1,6367e-03)	1,47e-01*** (3,1433e-03)
Post.Sec and college	5,05e-01*** (1,1759e-03)	3,42e-01*** (5,8848e-03)	4,38e-01*** (2,6157e-03)	4,12e-01*** (5,2695e-03)

Residuals S.E.	0.3280	0.2911	0.2970	0.2993
AdjustedR-squared	0.5865	0.4303	0.4833	0.5230
P-value(F)	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

Source: Author's own calculations using R software and *Quadros de Pessoal* dataset.
Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ^{2} is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \dots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model specified does better than the intercept to account for the regressand variability.

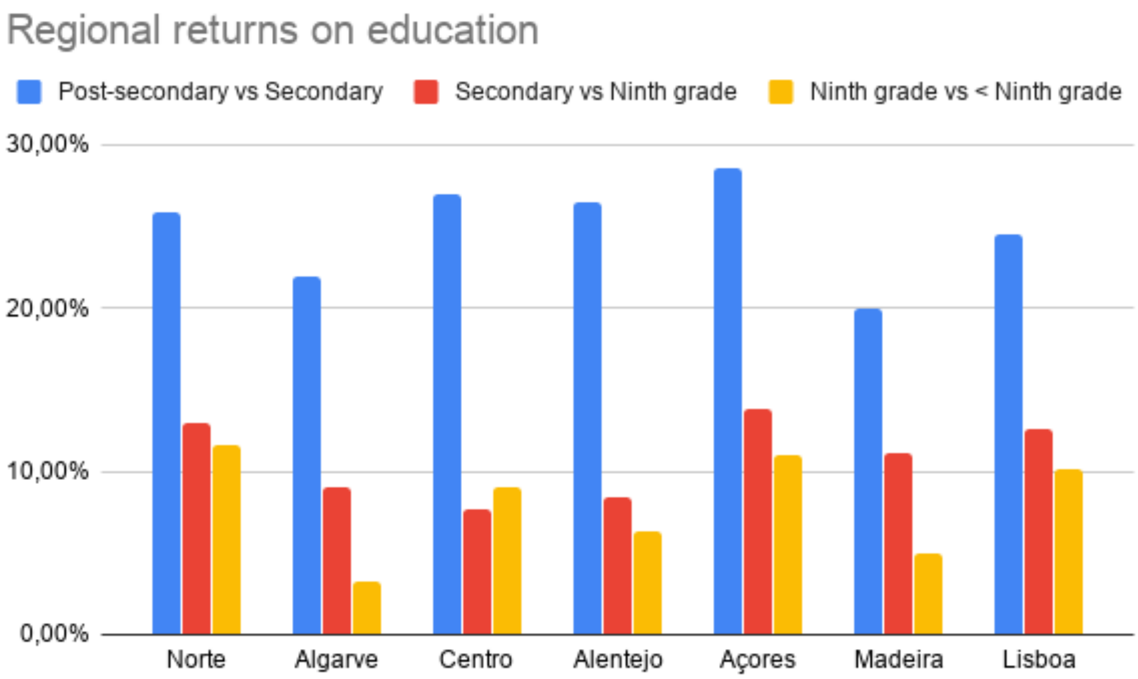
Table A7. Mincerian regression controlling for economic activity, work position and gender using regional data for 2015 using OLS with robust errors

	(AÇORES)	(MADEIRA)	(LISBOA)
(Intercept)	9,05e*** (1,4189e-02)	1,00e*** (11,7692e-02)	1,10e*** (5,7041e-03)
Exp	2,12e-02*** (5,5477e-04)	1,78e-02*** (5,5461e-04)	1,97e-02*** (1,2987e-04)
Exp ^{2}	-2,90e-04*** (9,7597e-06)	-2,45e-04*** (9,5202e-05)	-2,95e-04*** (2,4012e-06)
Tenure	1,22e-02*** (2,3754e-04)	9,69e-03*** (2,2429e-04)	1,52e-02*** (5,9822e-05)
Ninth_grade	1,10e-01*** (4,0980e-03)	5,03e-02*** (4,1501e-03)	1,01e-01*** (5,6994e-03)
Secondary	2,48e-01*** (5,3811e-03)	1,61e-01*** (5,2716e-03)	2,27e-01*** (1,3981e-03)
Post.Sec and college	5,34e-01*** (9,4111e-03)	3,60e-01*** (9,347e-03)	4,72e-01*** (1,9378e-03)

Residuals S.E.	0.2675	0.2987	0.3420
AdjustedR-squared	0.6715	0.6211	0.6428
P-value(F)	< 2.2e-16	< 2.2e-16	< 2.2e-16

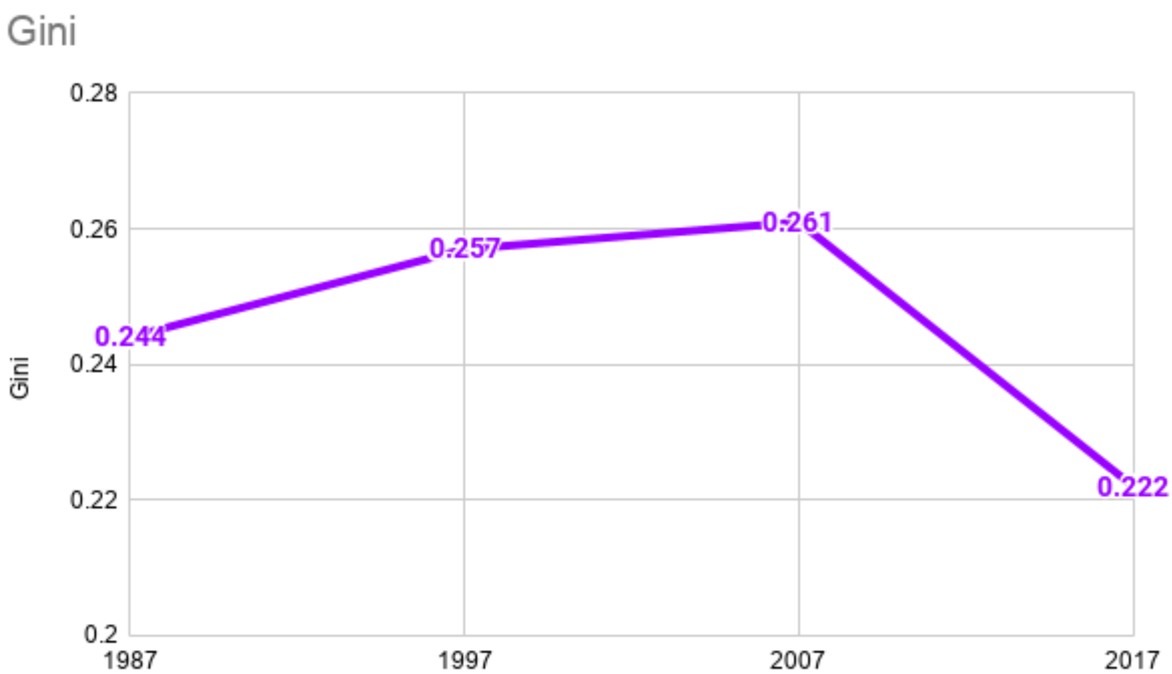
Source: Author's own calculations using R software and *Quadros de Pessoal* dataset. Notes: standard error in parenthesis. ***; **; * indicate statistical significance at 1%, 5% and 10% levels, respectively. Regressors: Exp stands for the years of experience of the worker, Exp ^{2} is the years of experience raised to the second power; tenure is the number of years spent by the worker at the current company; Ninth_grade, Secondary, Post_seconday and College or more are dummy variables that assume value 1 if the worker has, respectively, nine years, 12 years, 13 or 14 years and 15 or more years of formal schooling. The null hypothesis of the F-test is: $\beta_1 = \beta_2 = \beta_3 \dots \beta_n = 0$. A low p-value for the F-test means the joint rejection of the null hypothesis and the model specified does better than the intercept to account for the regressand variability.

Figure A2. Regional wage premiums for 2015



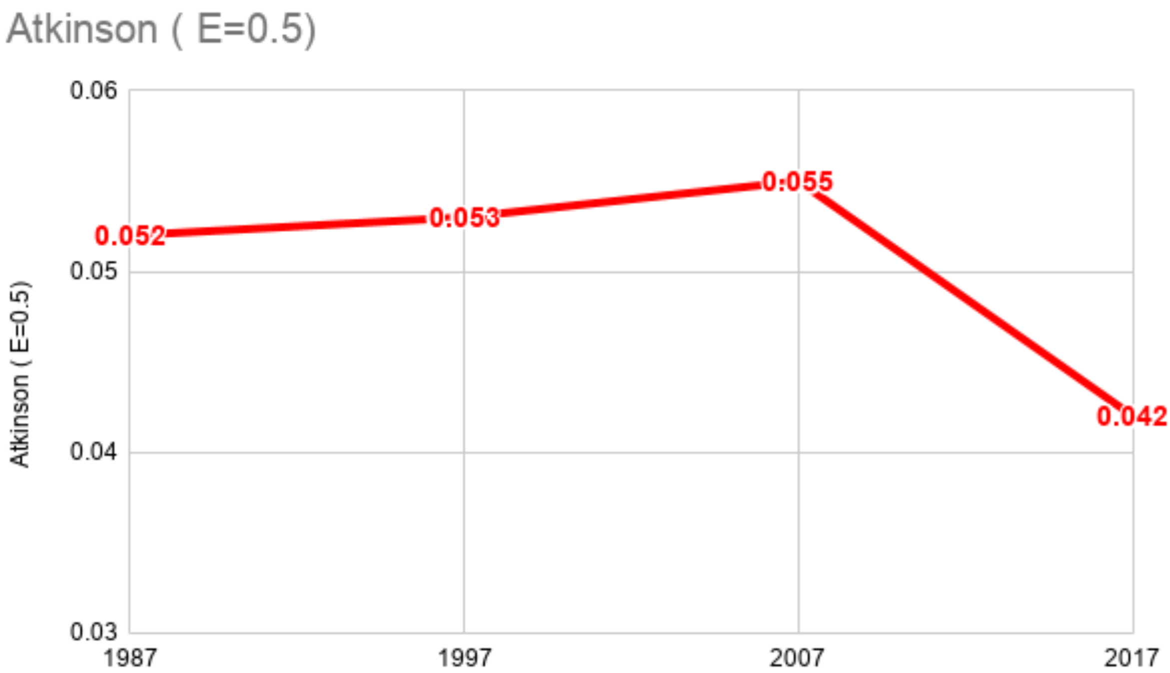
Source: elaborated by the author using *Quadros de Pessoal* dataset.

Figure A3. Human Capital inequality measured by the Gini coefficient



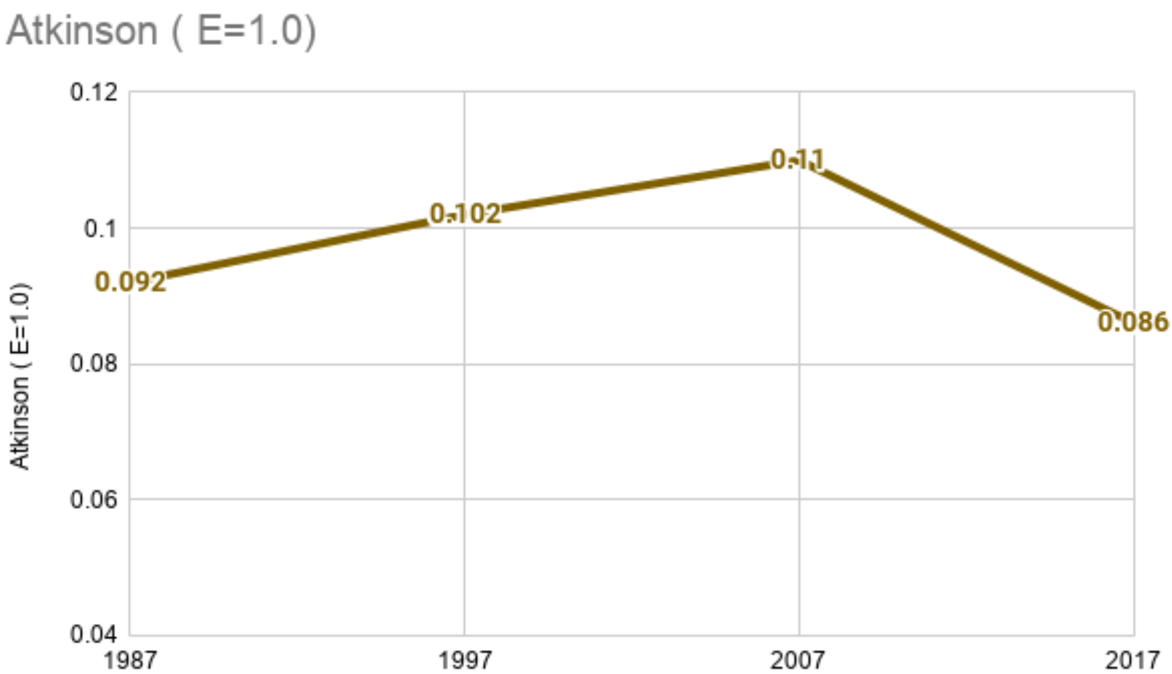
Source: elaborated by the author using *Quadros de Pessoal* dataset

Figure A4. Human Capital inequality measured by the Atkinson index ($\varepsilon = 0.5$)



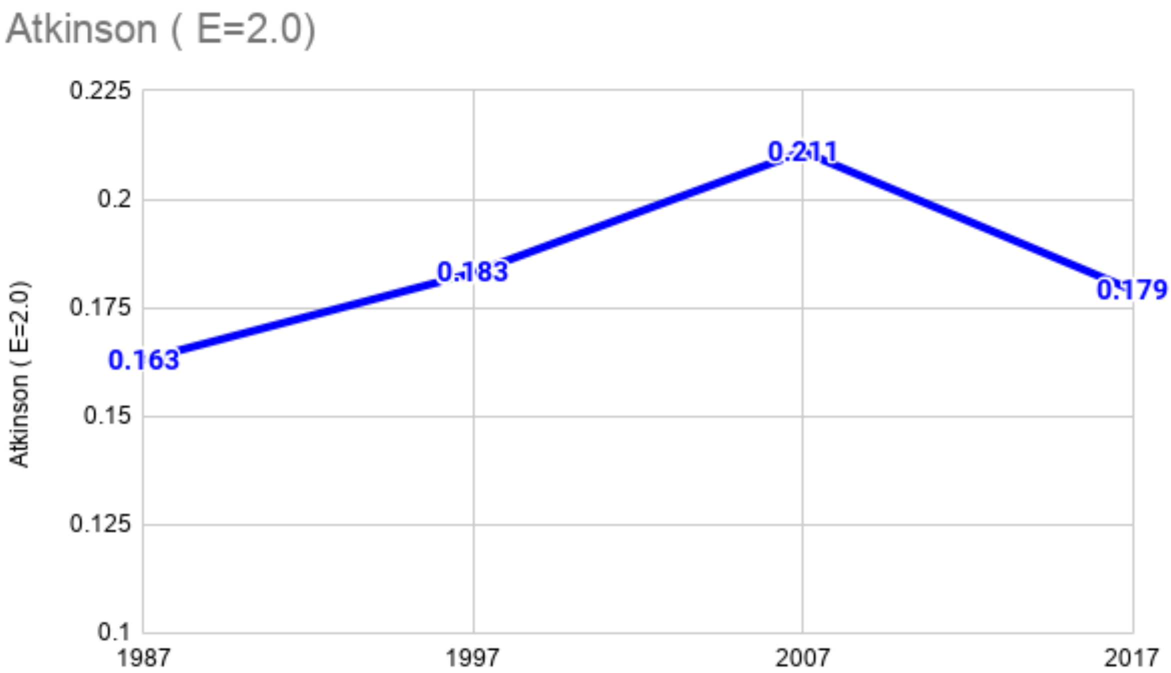
Source: elaborated by the author using *Quadros de Pessoal* dataset

Figure A5. Human Capital inequality measured by the Atkinson index ($\varepsilon = 1.0$)



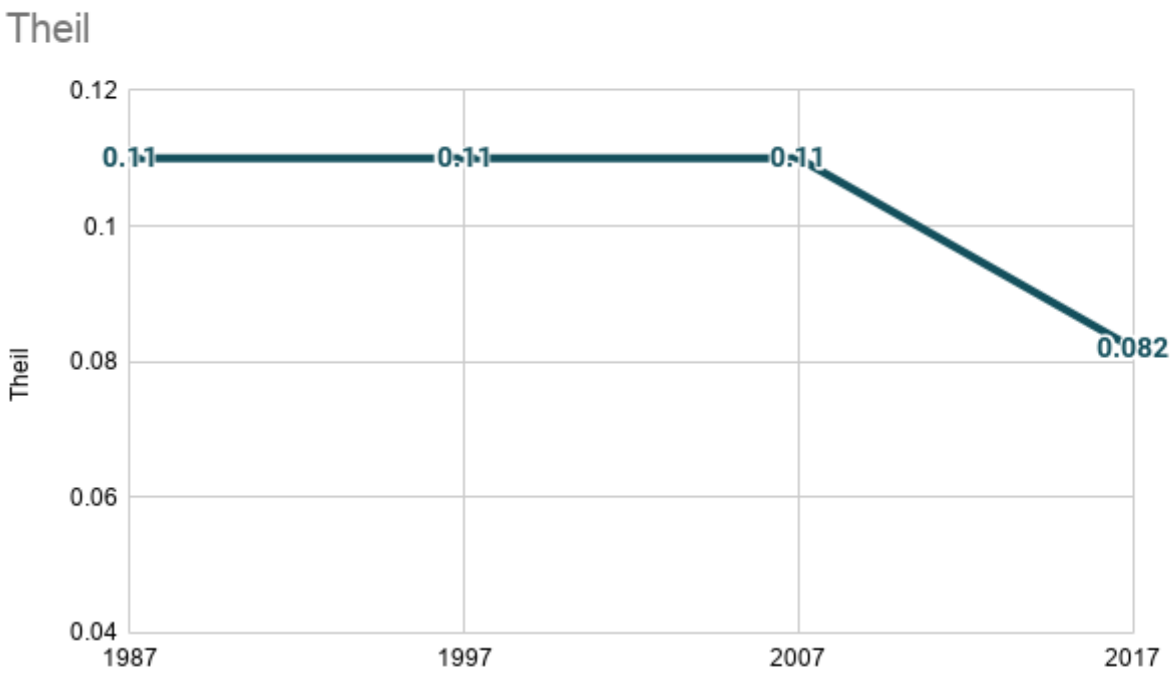
Source: elaborated by the author using *Quadros de Pessoal* dataset

Figure A6. Human Capital inequality measured by the Atkinson index ($\varepsilon = 2.0$)



Source: elaborated by the author using *Quadros de Pessoal* dataset

Figure A7. Human Capital inequality measured by the Theil's T index



Source: elaborated by the author using *Quadros de Pessoal* dataset