Wage Inequality: Trends and Drivers in Portugal

Francisco Espiga

Dissertation written under the supervision of Joana Silva

Dissertation submitted in partial fulfilment of requirements for the MSc in Economics, Major in Public Policy and Regulation, at the Universidade Católica Portuguesa, 22nd of April 2020

Wage Inequality: Trends and Drivers in Portugal

Francisco Espiga

Abstract

This thesis examines the evolution of wage inequality and its potential drivers, using harmonized household surveys and longitudinal matched employer-employee data from Portugal. Wage inequality was relatively stable until 2013 and has fallen since then. Changes in the variance of wages across workers with different skills and sector-occupations contributed to the change in wage inequality. In particular, the reduction in the education and experience premiums and the compression of wages across sector-occupations, played a major role over the past years. Nevertheless, a significant part of the change in inequality occurred among workers with similar skills and sector-occupations. Evidence based on additive worker and firm fixed effects models shows that heterogeneity across firms' pay premiums reduced wage inequality. In contrast, increased heterogeneity across workers contributed to the relative stability of wage inequality during 2004-13. The evidence also suggests that wage dispersion within firms was relatively constant over time, and most of the registered changes were associated with changes in the wage dispersion between firms. Finally, the evolution of wage inequality differed across business cycles. During the crisis period of 2009-13, wage decreases across all wage percentiles led to a relatively stable inequality trajectory. During the non-crisis period, wage improvements among low-wage sectors and occupations and deterioration among high-wage sectors and occupations led to a gradual reduction in wage inequality.

Keywords: wage inequality, skill groups, education and experience premium, sector-occupation groups, fixed effects, worker, firm, business cycles

Desigualdade salarial: tendência e vetores em Portugal

Francisco Espiga

Resumo

Esta tese examina a evolução da desigualdade salarial e potenciais fatores, usando dados harmonizados de inquéritos às famílias e dados longitudinais empresa-trabalhador em Portugal. A desigualdade salarial permaneceu relativamente estável até 2013 e caiu a partir daí. Mudanças na variância dos salários entre trabalhadores com diferentes aptidões que trabalham em diferentes sectores-ocupações contribuíram para a mudança na desigualdade. Em particular, a redução dos prémios de educação e experiência e compressão dos salários entre sectores-ocupações, desempenharam um papel importante nos últimos anos. No entanto, uma parte significativa da mudança na desigualdade ocorreu entre trabalhadores que têm aptidões e que trabalham em sectores-ocupações semelhantes. Os resultados baseados em modelos aditivos de efeitos fixos de trabalhador e empresa mostram que a heterogeneidade entre prémios pagos pelas empresas reduziu a desigualdade salarial. Contudo, o aumento da heterogeneidade entre trabalhadores atuou no sentido contrário, contribuindo para a relativa estabilidade da desigualdade salarial durante o período 2004-2013. Os resultados também sugerem que a dispersão salarial dentro das empresas foi relativamente constante ao longo do tempo e a maioria das mudanças registadas estiveram associadas a mudanças na dispersão salarial entre empresas. Finalmente, a evolução da desigualdade salarial diferiu entre ciclos económicos. No período de recessão/crise de 2009-13, as reduções salariais nos diferentes percentis salariais levaram a uma trajetória de desigualdade relativamente estável. Durante o período de não crise, as melhorias salariais entre os setores e ocupações de baixos salários e a deterioração entre setores e ocupações com altos salários levaram a uma redução gradual da desigualdade salarial.

Palavras-chave: desigualdade salarial, grupos de capacidade, retribuições de educação e de experiência, grupos de setores-ocupações, efeitos fixos, trabalhador, empresa, ciclos económicos

Acknowledgements

I would like to start by thanking my dissertation supervisor Joana Silva for all the guidance and help over the past few months. This dissertation wouldn't have been possible without the technical expertise and support of Professor Joana, that always motivated me to improve my dissertation. I appreciate all the availability for regular meetings in the last months and trust in my work.

I would also like to thank Instituto Nacional de Estatística (INE) for providing me access to fundamental databases (*Quadros de Pessoal* and *Inquérito às Condições de Vida e Rendimento*) used in this study, as well as, for all technical support when necessary. I am also grateful to professors Miguel Portela and Pedro Raposo for helping me with the cleaning of the datasets used in this dissertation, in particular with the coding used on Stata.

I am also thankful to Martim Leitão for the precious help in my thesis's language revision.

A special thanks to Catarina David and all my colleagues from the Applied Economics dissertation seminar for all the mutual help and support with the dissertation throughout this last semester.

Finally, I would like to thank my family, friends, university and all people that directly and indirectly contributed for my success during this last years at the university.

Contents

List of Figures

List of Tables

1. Introduction

A large body of research studies the increase in inequality and its drivers. Although this trend has marked many developed countries (Acemoglu and Autor 2011; Atkinson 2008) and developing countries (Ge and Yang 2014; Lee and Wie 2017; Lee and Wie 2015), inequality is not increasing everywhere. Recent evidence from Portugal suggests that household income inequality has decreased since 2004 (Alves, Cardoso, and Monteiro 2020; Arnold and Rodrigues 2015). However, this trend has become less pronounced since the financial and sovereign debt crisis.

This thesis examines the evolution and drivers of wage inequality, which is the main component of income inequality[1], in Portugal from 2004 to 2018. It focuses on the contribution of worker characteristics, such as education and experience, as well as firm heterogeneity. Most of the literature emphasizes the role of "secular forces", such as technological change and trade (e.g. Autor, Katz, and Krueger 1998; Acemoglu and Autor 2011; Krueger 2012; Autor and Dorn 2013; Feenstra and Hanson 1999; Autor, Dorn, and Hanson 2013; Autor et al. 2014; Pierce and Schott 2016). This thesis highlights the role of "cycle dependent forces", by examining the evolution of wage inequality in periods of slowdown and crisis, versus periods of economic growth.

The thesis uses a rich combination of (harmonized) household survey and longitudinal administrative employer-employee data for Portugal. It has four main parts. The first part documents the pattern of wage inequality in Portugal, comparing it with other Organisation for Economic Cooperation and Development (OECD) countries using similar household survey data. The second part uses richer administrative data from Portugal to analyze labor earnings inequality by workers' key socio-economic characteristics, computing education and experience premiums, and disentangling the contribution of changes in between- and within-group inequality to changes in inequality. Then, the analysis is extended by including sector-occupation groups, highlighting how wage differences across sectors and occupations contributed to wage inequality changes. In the third part, this analysis is complemented by a study of the role of firm heterogeneity, using longitudinal matched employer-employee administrative data. The study decomposes changes in wage inequality over time into firm and

[1] Wages in Portugal represent approximately 65% of total household income in 2017 (computed using data from PORDATA and INE).

worker fixed-effects, a model that was first proposed by Abowd, Kramarz, and Margolis (AKM, 1999). It disentangles the contributions of unobserved worker and firm heterogeneity to changes in wage inequality and emphasizes the evolution and contribution of wage dispersion within and between firms to inequality in the Portugal. The fourth part of the thesis relates the business cycle to wage inequality in the country through differential wage dynamics in crisis and non-crisis periods.

The analysis yields four main empirical results. First, in contrast with the reduction in household income inequality, wage inequality in Portugal remained relatively stable until 2013, with a small reduction between 2009 and 2011 (as in Portugal, Raposo, and Reis 2018). But between 2013 and 2018, wage inequality decreased. This reduction resulted from a much faster wage growth at the bottom 5th and 25th percentile of the wage distribution than at the median, 75th and 95th percentile. This trend is particularly relevant, given that the international comparisons developed in this thesis show that the level of wage inequality in Portugal is above the OECD average, although it is below the levels in the United States and Latin America. Interestingly, compared with older people, younger individuals contributed more to the downward trend in wage inequality in Portugal.

Second, the education premium has fallen over time, contributing to less wage inequality. The premium associated with firm tenure (time working at the same firm) also fell throughout the whole period. In contrast, the labor market experience premium has been almost stagnant until the past few years. Decomposing the contribution to the overall change in wage inequality of between- and within-skill-groups changes in wage inequality, defined in terms of education and experience, I find that the importance of the between-skill-groups component surpassed that of the within-group component and was larger after 2013. This contribution increased significantly when workers' sector-occupation groups are included. The changes in the variance between sector-occupation groups played a larger role in inequality dynamics after 2013, and their contribution to the overall change in wage inequality surpasses the one from the within component.

Third, firms had a key role in the recent wage inequality dynamics in Portugal. Using AKM models to decompose wage variance, heterogeneity across workers explains between 60% and 70% of the overall inequality, while firm heterogeneity explains between 22% and 28%. This result is consistent with the literature, even if there might be some differences on the specification used (see, for example, Portugal, Raposo and Reis (2018) for evidence for Portugal, and Alvarez et al. (2018) and Card, Heining, and Kline (2013) for evidence for Brazil and Germany, respectively). However, changes in firm heterogeneity has the largest contribution to the reduction of overall inequality during the period under analysis. Between 2004 and 2013, the small reduction in wage inequality (mainly in 2009 and 2010) was due to the reduction in differences across firms and the association between highly remunerated workers and high-paying firms. However, heterogeneity across workers increased over the same period and contributed to a more stable trend in inequality. Between 2009 and 2017, wage inequality also decreased due to the reduction in the differences across workers and firms, contributing to the downward trend after 2013. When decomposing inequality in the wage variance between- and within-firm, the within-firm dispersion remained with a smaller contribution and relatively stable, compared with the between-component, and exhibited a smaller contribution to the change in the overall wage inequality. These results are consistent with Messina and Silva (2019) for Latin America and Song et al. (2019) for the United States.

Fourth, during the crisis period (2009-13), all wages fell, both in high- and low-earnings sector-occupations, which explains the relative stability of inequality. During the period that followed (2014-17), there was a combined effect of wage increases among sector-occupations earning less and wage decreases among sector-occupations earning more, resulting in lower inequality. These results differ from those of Autor, Katz, and Kearney (2008) where United States' wage inequality was shown to be increasing.

This thesis examines wage inequality in Portugal, contributing to three strands of previous literature. First, it contributes to a large literature studying the role of workers' characteristics, such as education, experience, or occupation, in wage inequality dynamics. Most of this literature focuses on the United States, showing that the increase in wage inequality registered between the 1960's and the 2000's was strongly associated with higher differentials between gender-experience-education groups, and between college and high school graduates (e.g., Autor, Katz, and Kearney 2008; Goldin and Katz 2008; Acemoglu and Autor 2011). Latin America has recently received renewed attention, as wage inequality has fallen significantly since 2002 (except in Costa Rica). Recent papers find that two-thirds of the decline in the wage inequality in the region has occurred within skill groups. Although changes in the sectoral, occupational, and formal/informal composition of the workforce matter, they do not fully explain the reduction in within-skill variance (e.g. Messina and Silva 2019). This thesis contributes to this literature by re-examining these questions for Portugal and decomposing the contribution (and relative importance) of changes in both between- and within-group wage inequality to the overall changes in wage inequality in a country with subperiods with markedly different inequality dynamics, and using datasets that cover the whole economy.

Second, the thesis contributes to a growing literature that highlights the role of firms in wage inequality dynamics, using linked employer-employee datasets. Card, Heining, and Kline (2013) document an increase in wage inequality from 1985 to 2009 in West Germany. They show that the key factors behind this trend were greater dispersion across workers and firms and a higher association between high-paying firms and highly paid workers. Studies of the United States between 1978 and 2013 have documented an increase in wage inequality, where the main contributor was the heterogeneity across workers, followed by the covariance of worker and firm effects. In these cases, the heterogeneity across firms decreased. However, when inequality is decomposed into differences between and within firms in the United States, both these terms contributed to the increase in wage inequality (e.g., Song et al. 2019). Recent studies of Latin America also confirm the decisive role played by heterogeneity across firms in the reduction of wage inequality, even if the majority of this inequality is still associated with heterogeneity across workers (e.g., Alvarez et al. 2018; Messina and Silva 2019). A recent paper on Central and Eastern European countries, which also decomposes wage inequality into between- and within-firm components, shows that dispersion across firms was the main contributor to the decrease in wage inequality in the majority of these countries, although dispersion within firms also diminished between 2006 and 2014 (e.g., Magda, Gromadzki, and Moriconi 2019). For Portugal, using high-dimensional fixed effects models, recent papers show that heterogeneity between workers contributed the majority towards wage inequality throughout 1990's, followed by heterogeneity across firms. The relatively stable worker and firm components contributed to a period of relative stagnation of wage inequality after 1992 (despite a decrease in 2009/2010), even though the association between well-paid workers and high-paying firms followed a downward trend (e.g., Torres et al. 2013; Portugal, Raposo, and Reis 2018). This thesis contributes to this literature by re-examining these issues, covering the recent years where changes in the overall inequality were more significant.

Third, this thesis contributed to the literature that relates the business cycle with wage inequality. It adapts the approach of Autor, Katz, and Kearney (2008). These authors analysed wage variations along occupational skill percentiles in the United States in different subperiods, studying wage growth for low- and high-wage workers. In this thesis, through the study of changes in wages across sector-occupation wage percentile in crisis and non-crisis subperiods, it is possible to characterize wage variations across the business cycle and verify who are the groups (workers in poorer or richer sector-occupations) contributing the most to the wage inequality pattern.

The thesis proceeds as follows. Section 2 presents the empirical methodology. Section 3 describes the data used in the analysis. Section 4 provides the results on the study of wage inequality patterns and drivers. Section 4.1 describes wage inequality patterns in Portugal since 2004, compares them with other OECD countries, and analyses wage inequality dynamics across demographic groups. Section 4.2 shows the evolution of education and experience premiums through time and the related changes on labor supply. The analysis is also complemented by a decomposition of wage inequality changes into changes in the between- and within-skill- (and sector-occupation) components. Section 4.3 presents new evidence on the role of worker and firm heterogeneity for changes in wage inequality, and decomposes the overall change in wage dispersion in its within- and between-firm components. Section 5 examines wage inequality dynamics during periods of crisis and of positive economic growth in Portugal. Finally, section 6 concludes.

2. Empirical methodology

This thesis has three main parts that use distinct empirical methodologies. The initial part of the thesis documents wage inequality trends for Portugal since 2004 and compares them with those in other OECD countries[2]. It assesses trends in three main wage inequality indexes: the Gini Coefficient, the variance of earnings and the log 90/10 differential. This part of the thesis uses year-to-year cross-sectional data for Portugal. For comparability, Portuguese information on wages was harmonized using the same methodology as the one applied to other countries (see detail below).

The second part of the thesis uses the richness of the Portuguese administrative data to compare “composition” adjusted education and experience premiums, studying the pattern of the evolution of these premiums over time, disregarding changes on each group composition. The

[2] Including Germany, Greece, Sweden, Hungary, United Kingdom, United States, Japan and Colombia.

methodology adopted to adjust for changes in composition is based on Autor, Katz and Kearney (2008), Acemoglu and Autor (2011), Centeno and Novo (2014) and Messina and Silva (2019). It includes four steps. First, worker data in each year is allocated between 24 education-experience-gender groups (skill groups). The skill groups are obtained from three education categories (less than high school education, high school education, college education or more), four experience categories (0-9, 10-19, 20-29, 30 or more years of experience) and two genders[3]. Second, composition-adjusted predicted log of real hourly wages at each skill group and year are estimated from regressing, for each gender and year separately, the log of real hourly wages on the interaction terms of education and experience categories. Third, weighted averages of predicted log of real hourly wages are computed for broader education or experience groups, using mean employment share of each skill group from 2004 to 2017 as a fixed weight, obtaining the expected wages of a worker with some education or experience skill for each year[4]. Fourth, education or experience premiums (log relative real hourly wages between different groups) are obtained by differentiating average predicted log of real hourly wages across education or experience groups in each year.

Afterwards, the previous analysis was complemented by further disentangling the factors behind overall wage inequality trends. It starts with a decomposition of the overall variance of log wages on its between- and within-group components to quantify their relative importance. Skill groups are still defined by education, experience and gender categories as before. Following Messina and Silva (2019), the within-skill component of overall variance is measured by estimating a standard Mincer regression having skill variables and all their interactions as regressors. Specifically, the estimated regression is:

$$w_{it} = x'_{it}\beta_t + \varepsilon_{it} \tag{1}$$

where w_{it} stands for log of real hourly wages of worker *i* in year *t*, x'_{it} is the vector of all interactions between education, experience and gender categories for worker *i* in year *t*, β_t is the vector of the return to skills in year *t* and ε_{it} is the error term/residual of worker *i* in year *t*.

[3] Education is divided in three categories based on worker's highest education level completed. Worker's observations without information on education were eliminated. Workers that completed technical courses are considered as having completed high school education. Experience is determined through worker's age minus 16 (workers completed less than high school level) or 18 (workers completed high school level) or 21 (workers completed college level). Worker's observations with negative values on experience were eliminated.
[4] Other alternative fixed weights can be used in the analysis. Employment share of each group in year 2017 was also considered as an alternative weight, demonstrating similar results to the ones presented.

The variance between skill groups, after computing the variance of log of real hourly wages in each year, was given by the overall variance minus the variance within skill groups. Due to the linearity of this decomposition, it also holds for changes in the variance, the variable of interest.

To further decompose the changes within and between groups, the sector-occupation pay differentials dimension was considered and equation (1) was extended as follows:

$$w_{it} = x'_{it}\beta_t + \sigma_{SO(i,t)} + \varepsilon_{it} \tag{2}$$

where $\sigma_{SO(i,t)}$ stands for sector-occupation of worker i at year t. After estimating regression (2) in each year, the total wage variance was decomposed in variances among sector-occupations and skill groups, twice the covariance between skill and sector-occupation groups and variance of residual component. The decomposition can also be applied to the changes on variance of log real hourly wages.

The third part of the analysis, examines the role of firms in which workers are employed, a dimension of wage dispersion that the previous analysis did not consider. Workers of different firms may have differences on their earnings even if skills and sector-occupations are identical. Thus, firms become an important dimension of wage inequality, and studying their heterogeneity may shed light on inequality patterns. This thesis decomposes the wage variance on the worker and firm components to evaluate the contribution of firm's heterogeneity to wage inequality using high dimensional fixed effects model (the AKM model, first proposed by Abowd, Kramarz and Margolis 1999) by subperiods (2004-2009, 2009-2013 and 2013-2017), following Card, Heining and Kline (2013); Alvarez et al. (2018), Messina and Silva (2019) and Song et al. (2019) approach. The AKM regression model estimated in each subperiod is:

$$w_{it} = \omega_i + \delta_{F(i,t)} + \tau_t + \varepsilon_{it} \tag{3}$$

where w_{it} is the log of real hourly wages of worker i at year t and it is given by the sum of ω_i, $\delta_{F(i,t)}$, τ_t and ε_{it} standing for the worker component, firm component, year effect and residual component, respectively, assuming that residual respect a strict exogeneity condition regarding the other components. Due to the use of longitudinal matched employee-employer data, allowing to follow workers through employers and time, and using Correia (2014) estimation method for high dimensional fixed effects model, it was possible to obtain the workers, firms and year fixed effects in each subperiod. This specification allows to analyze, separately, the earnings for different fixed observable or unobservable worker's characteristics (for example worker's ability) remunerated equally across firms from the persistent firm's pay premiums (worker's remuneration at each firm associated, for example, with rent-sharing)[5]. The error term accounts for temporary wage levels of workers and the fluctuations that might occur. After estimating the fixed effects on the previous AKM model, the wage variance in each subperiod was linearly decomposed in:

$$Var(w_{it}) = Var(\omega_i) + Var(\delta_{F(i,t)}) + Var(\tau_t) + 2 * Cov(\omega_i, \delta_{F(i,t)}) + 2 * Cov(\omega_i, \tau_t) + 2 * Cov(\delta_{F(i,t)}, \tau_t) + Var(\varepsilon_{it}) \quad (4)$$

where $Var(w_{it})$ stands for the variance of log real hourly wages, $Var(\omega_i)$, $Var(\delta_{F(i,t)})$, $Var(\tau_t)$ and $Var(\varepsilon_{it})$ are the variance of worker, firm, year and residual component, respectively, $Cov(\omega_i, \delta_{F(i,t)})$ is the covariance of worker and firm component, $Cov(\omega_i, \tau_t)$ is the covariance of worker and year component and $Cov(\delta_{F(i,t)}, \tau_t)$ is the covariance of firm and year component. Based on the previous interpretation of the worker and firm component, the variances represent wage heterogeneity across workers related to their fixed characteristics (observable or unobservable) and wage heterogeneity across firms from their distinct pay premiums to employees.

3. Data description

The analysis in this paper draws on two main datasets:

(i) *Inquérito às Condições de Vida e Rendimento* (ICOR), which is collected by the Portuguese Institute of Statistics. ICOR is the Portuguese annual household survey and contains information on earnings, demographic characteristics (gender, age, schooling) and job characteristics (job situation, self-definition of working conditions, work experience, occupation and hours of work) of individuals and their families (when applied), besides the information and data on health and living conditions. The ICOR data are available for the period 2004-2018, with close to 29400 individuals and 13700 families covered in 2018. Since ICOR didn't exist before 2004, the first year studied is 2004. This thesis restricts attention to full-time dependent workers aged between 18 and 65 years old (working age population). It

[5] As in Alvarez et al. (2018) and Messina and Silva (2019), time-varying observable characteristics of workers are not introduced in the model, avoiding the need to consider their effects in small subperiods when changes occur in the within component.

uses real gross hourly wages in the main activity of the worker. Sample weights were applied in all computations and estimations. Appendix A1.1 provides more details on the ICOR database and on the treatment of the data used.

(ii) *Quadros de Pessoal (*QP), which is collected by the Portuguese Ministry of Employment. It covers virtually all workers and firms in the Portuguese private sector with more than one worker, having around 200,000 firms and more than 2 million workers each year over the 2004-2017 period. It provides comprehensive information on workers' demographic characteristics (age, gender, schooling), job characteristics (occupational group, professional category, wage, hours worked and firm tenure), along with employing firm identifier codes. Firm-level characteristics include sales, number of employees, equity, percentage of foreign capital, geographical location and date of creation, along with industry code. As the Portuguese classification of firm's economic sectors (*Classificação Portuguesa das Atividades Económicas* - CAE) has been revised in 2007 to match the NACE Rev.2, a concordance was needed and the analysis since 2007 considers 20 sectors according to the one-digit level classification. In the case of the occupation classification, it was revised in 1994 (*Classificação Nacional de Profissões* - CNP) and 2010 (*Classificação Portuguesa de Profissões* - CPP). A concordance was also needed and the analysis done since 2010 considers 11 major groups of CPP/2010. In QP, analysis is restricted to full-time dependent workers aged between 18 and 65 years old (working-age population) from 2004 to 2017. It is used real gross hourly wages on the main activity of the worker. More details and information on the QP database and data treatment in Appendix A1.2.

I supplement this data with information on wage inequality indicators for OECD countries computed using surveys comparable to ICOR.

4. Results

4.1. Wage inequality trends in Portugal and international comparisons

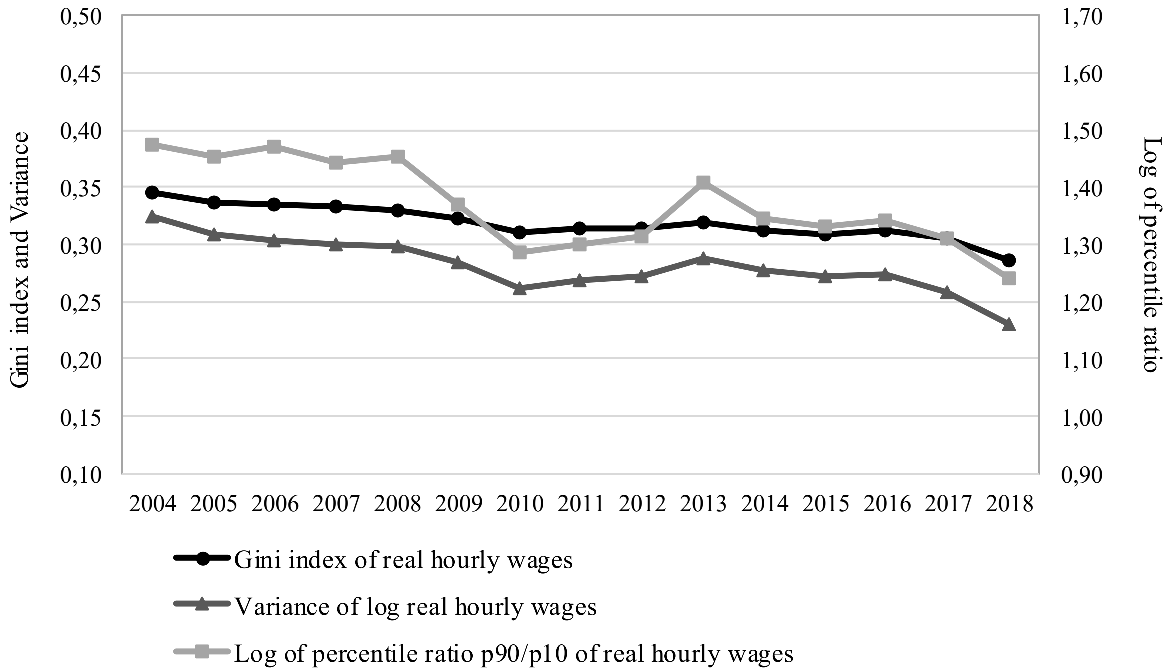
Figure 1 shows the evolution of wage inequality in Portugal since 2004 using ICOR data. It uses three different indicators: Gini index of real hourly wages[6], variance of log real hourly

[6] The Gini index measures wage's concentration, varying between zero (perfect equality) and one (perfect inequality).

wages and log of percentile ratio p90/p10 of real hourly wages[7]. These three indicators are important to check the robustness of the findings and serve as the basis for different parts of the remaining analysis. For example, the variance allows one to decompose the overall inequality in its components. Using log of real hourly wages on the variance has the advantage of not being sensitive to identical proportional changes in wages (Atkinson 1970).

A similar pattern emerges across all three indicators: inequality remained relatively stable until 2013, following a downward trend after this period. In particular, inequality had minor fluctuations before 2009, maintaining its level, followed by a decrease of almost 6%, 13% and 12% in the Gini index, variance and log percentile ratio, respectively, from 2008 to 2010. However, the slight increase in inequality between 2010 and 2013, contributes to the stability in wage inequality trend. After 2013, inequality fell around 10%, 20% and 12% when measured using the Gini index, variance and log percentile ratio, respectively, until 2018, presenting a downward trend.

Figure 1 – Evolution of wage inequality, Portugal

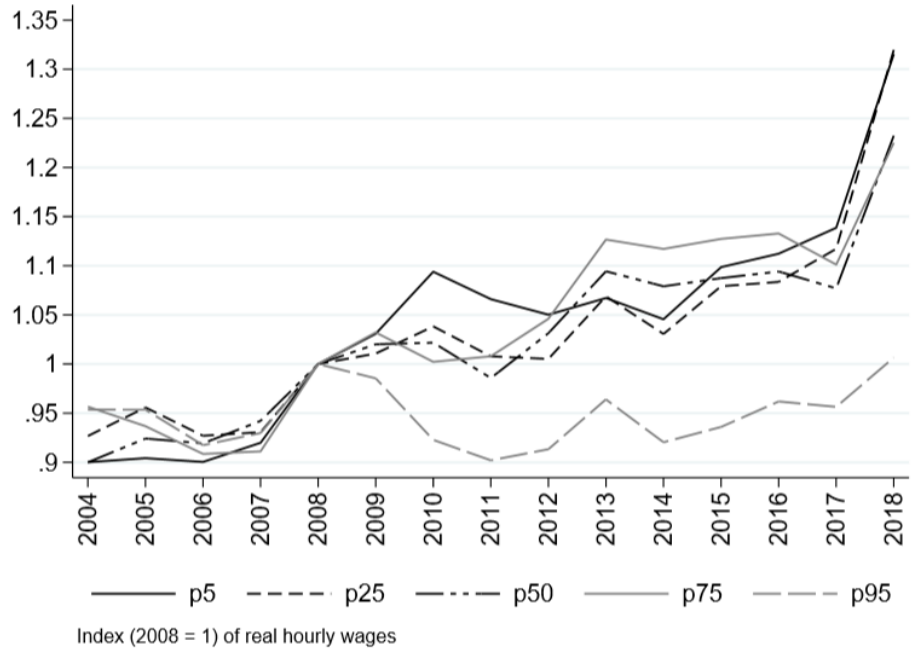


Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure presents three measures of wage inequality between 2004 and 2018: Gini index of real hourly wages, variance of log real hourly wages and log of percentile ratio p90/p10 of real hourly wages. Results are generated using cross-sectional sample weights. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

[7] The percentile ratio p90/p10 measures wage inequality, focusing on the wage differences between top and bottom deciles.

Figure 2 presents the 5th, 25th, 50th, 75th and 95th wage percentiles growth between 2004 and 2018[8]. Wage, as before, was defined as real hourly wages and 2008 was chosen as the index base year. Three findings emerge from this figure, justifying the evolution of wage inequality. First, the increase of almost 10% in the 5th percentile and decrease of around 7% in the 95th percentile between 2008 and 2010, contributed for the decrease in inequality in that period. Second, between 2011 and 2013 the deceleration of bottom percentiles growth and acceleration of top percentiles growth, contradicted the previous decrease on inequality. Third, the significant growth of more than 30% in bottom percentiles relative to 2008 in comparison with a lower growth at 50th, 75th and 95th percentiles, led to a reduction in wage inequality after 2013.

Figure 2 – Wage growth by percentile, Portugal



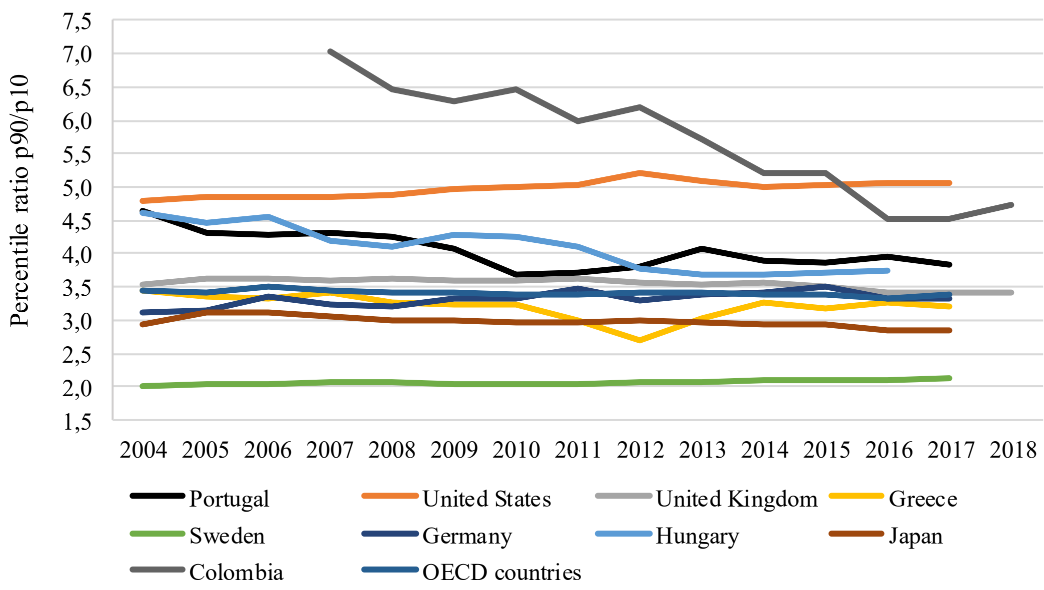
Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure presents the growth between 2004 and 2018 (index base at 2008) in the 5th, 25th, 50th, 75th and 95th wage percentiles. Results are generated using cross-sectional sample weights. Wages appear as gross real hourly wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Figure 3 provides a comparative analysis of wage inequality evolution since 2004 on several OECD countries, including Portugal, based on collected data from household or individual surveys for each country. In each country, wage inequality was measured for full-time dependent workers, using the percentile ratio p90/p10 of gross earnings. Inequality in Portugal had one of the highest levels (between 3,69 and 4,65) among the European countries and above

[8] Table A2.1 in Appendix A2 provides more information on the 5th, 25th, 50th, 75th and 95th wage percentiles in each year that were used to compute percentiles growth in figure 2.

OECD average level, but being surpassed by US and Latin American countries. Scandinavian countries, for example Sweden, have the lowest inequality levels in Europe (around a ratio of 2,0), while some Central and Eastern European countries, such as Hungary, are in the group of the most unequal countries in Europe. Portugal belongs to the latter group. In terms of changes in inequality, Portugal shows minor fluctuations in inequality in comparison with Latin America, even though there were some fluctuations and a decrease of around 18% in inequality between 2004 and 2017. The percentile ratio p90/p10 continuously decreased in Latin America countries, featuring for example a significant fall of almost 33% in Colombia between 2007 and 2018. In US between 2004 and 2017 and in Germany from 2004 to 2011, inequality grew almost 6% and 12%, respectively. Hungary, as an example of Eastern Europe countries, had fluctuations on wage inequality – this latter decreased by around 19% between 2004 and 2016. Therefore, since 2004, Portugal was one of the European countries in which inequality fluctuated the most, highlighting the relevance of studies on the evolution of inequality in Portugal.

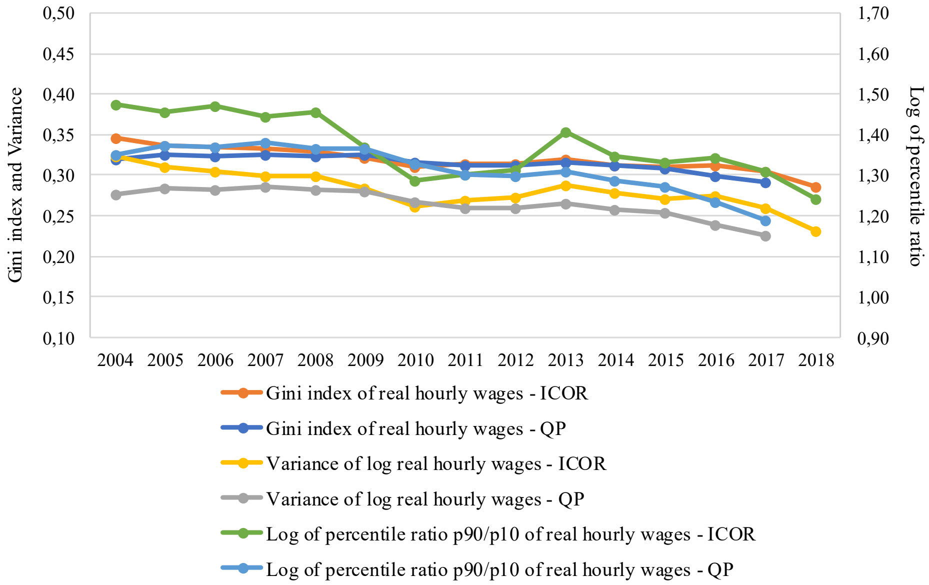
Figure 3 – Wage inequality in different OECD countries



Source: Data from OECD.Stat
Note: Figure presents the percentile ratio p90/p10 for Portugal, United States, United Kingdom, Greece, Sweden, Germany, Hungary, Japan, Colombia and OECD countries of gross wages for full-time dependent workers. Data from household or individual surveys done to each country.

Importantly, in order to make international comparisons, the current section studied initially inequality using ICOR datasets. However, as discussed in the methodological section, this thesis will use the richness of QP in subsequent parts. Unlike a survey, QP datasets have information on all dependent workers in Portugal, allowing to follow each worker over time (longitudinal data) and to identify firms where they work (linked employer-employee dataset), through unique identifiers of employees and employers. Hence, and to check the robustness of our findings to different data sources, Figure 4 plots wage inequality measures using ICOR and QP datasets. Results (in levels and trends) are remarkably similar. In particular, before 2009 inequality maintained its level, but from 2009 to 2011, the Gini index, variance and log percentile ratio decreased by 4%, 7% and 5%, respectively. However, they increase slightly in 2012 and 2013 contributing to the relative stability of wage inequality during this period. Importantly, these trends and levels are also consistent with previous papers analysing wage inequality in Portugal (Portugal, Raposo, and Reis 2018). As documented in this thesis using data from ICOR, data from QP confirms the downward trend in inequality registered between 2013 and 2017. Specifically, the Gini index, variance and log percentile ratio fell by around 8%, 15% and 9% respectively. Figure A2.1 in Appendix A2 presents a detailed analysis of wage growth by percentile, similar to figure 2, further confirming the consistency of wage inequality results across these two datasets[9].

Figure 4 – Evolution of wage inequality using different data sources



Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR) and *Quadros de Pessoal* (QP)

Note: Figure presents three measures of wage inequality: Gini index of real hourly wages, variance of log real hourly wages and log of percentile ratio p90/p10 of real hourly wages. Results in ICOR are generated using cross-sectional sample weights. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year and database were trimmed.

[9] Table A2.2 in Appendix A2 provides the 5th, 25th, 50th, 75th and 95th wage percentiles in each year that were used to compute percentiles growth in Figure A2.1.

The use of QP, allows us to further exploit a dimension of wage dynamics that Alves, Cardoso, and Monteiro (2020) suggest could have been of particular importance: age. Following Alves, Cardoso, and Monteiro (2020), Figure 5 presents the evolution of wage inequality for different age groups between 2004 and 2017. Two important findings emerge. First, wage inequality is higher among older age groups. This is in line with the stylized fact documented in recent studies of increasing earnings inequality throughout the life-cycle due to effects of cumulative shocks during worker's professional career (e.g. Alves, Cardoso, and Monteiro 2020). Second, in contrast with older age groups, inequality within younger age groups fell significantly during the period. Older age groups maintained a relatively stable wage inequality with a more continuous decrease only after 2013. In particular, between 2004 to 2017, the Gini index of new entrants fell by around 24% and variance by 41%, while that of 30-39 years old fell by around 17% and 31%, respectively. Before 2009, there was a slight increase of 4% and 6% in Gini index and variance, respectively, for the 50-65 years group. From 2009 to 2011, Gini index and variance for both 40-49 and 50-65 years old groups, decreased around 2% and 5%, respectively. After 2013, the Gini index of 40-49 and 50-65 years old groups decreased around 7% and 8% and the variance decreased 14% and 15%, respectively. Thus, older age groups were the main responsible for the stability of wage inequality between 2004-2013, even if they contributed for its reduction between 2009 and 2011. The common reduction across all age groups after 2013 contributed for the notorious downward trend in wage inequality during that period.

Figure 5 – Evolution of wage inequality by age groups

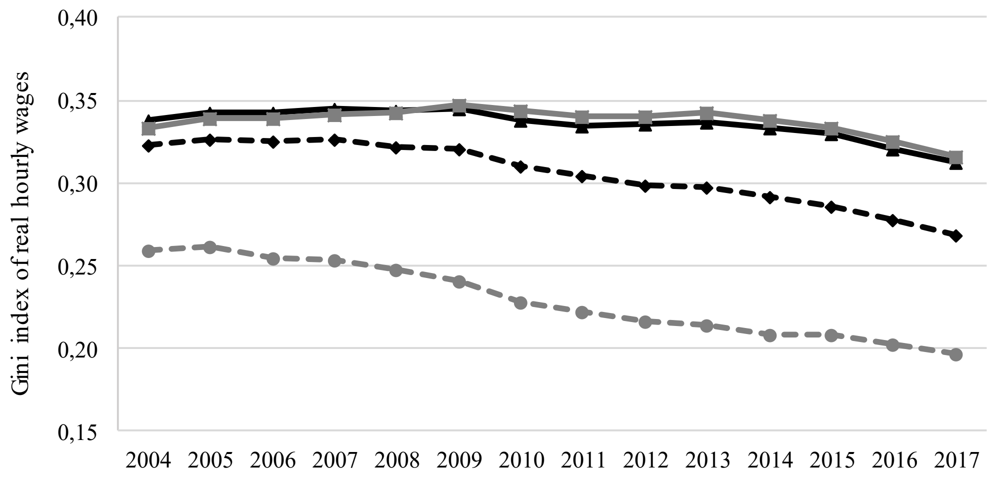
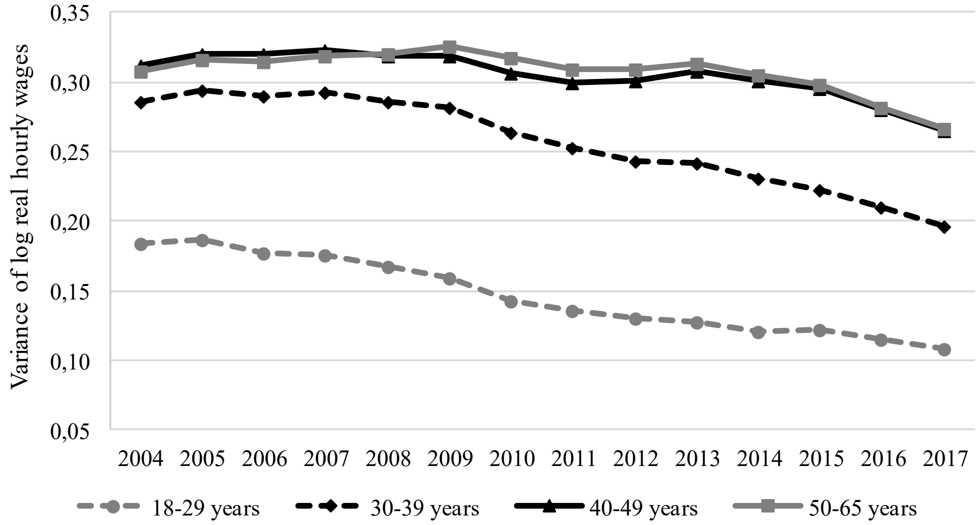


Figure 5 – Evolution of wage inequality by age groups (cont.)

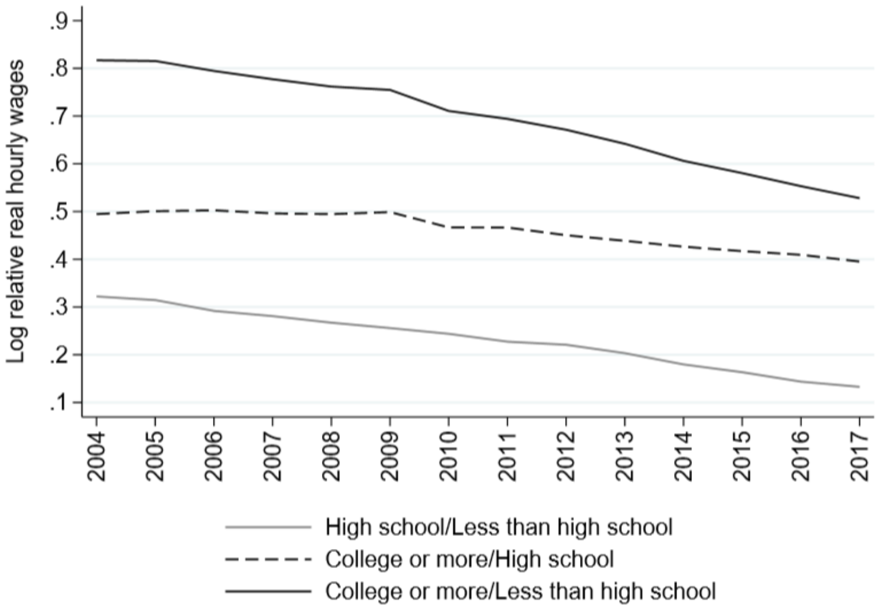


Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents, for each of the four age groups, two measures of wage inequality: Gini index of real hourly wages and variance of log real hourly wages. The four age groups used for the computations were: 18-29 years old (new entrants), 30-39 years old, 40-49 years old and 50-65 years old. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

4.2. Inequality between skill groups: analysis of education and experience premiums

While the previous section presents wage inequality trends, the present and the next sections analyze their potential drivers. In what follows, the focus lies on the contribution of composition-adjusted education and experience premiums. Figure 6 presents the evolution of education premiums between 2004 and 2017, showing a general downward trend. In particular, between 2004 and 2017, log relative real hourly wages between workers with a high school and less than high school level and between workers with college and less than high school education fell by around 0,19 log points (59%) and 0,29 log points (35%), respectively. The gradual reduction on log of real hourly wages for workers with only high school level or a college degree combined with a relative stability followed by an increase on earnings for worker with less than high school, contributed for the observed patterns. Wage gaps between college and high school educated workers remained stable until 2009. In the following years, when average earnings of college educated workers started to decrease faster than high school educated workers, wage gaps between college and high school educated workers fell by around 0,1 log points (20%). Given the observed decrease in wage differentials between workers with different education levels, wage inequality was expected to fall during the entire period, following a downward trend throughout, and not just since 2013. The wage inequality relative stability prior to 2013 suggests the existence of a countering effect from other simultaneous forces. Figure A2.2 in Appendix A2 replicates the analysis of composition-adjusted education premium using ICOR datasets, obtaining similar patterns.

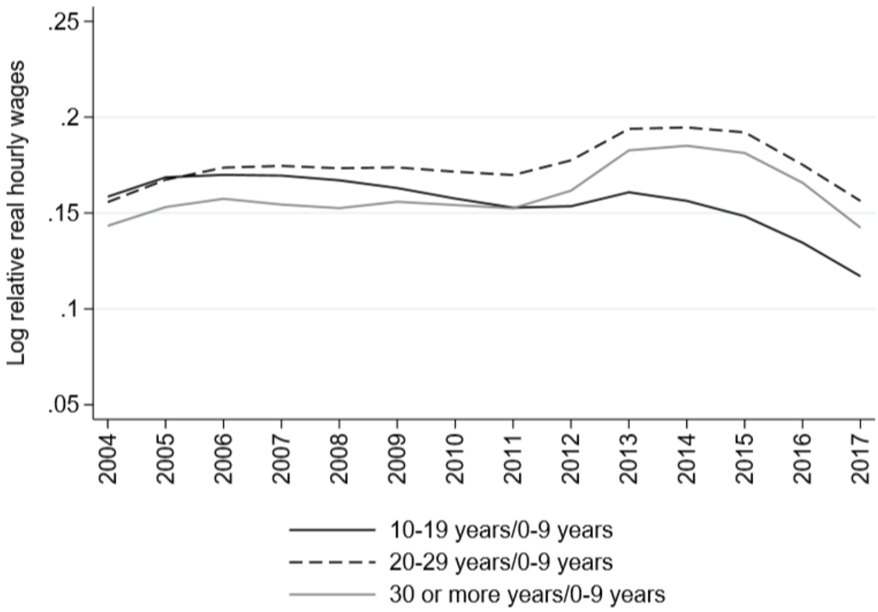
Figure 6 – Evolution of the composition-adjusted education premium



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents composition-adjusted education premium between 2004 and 2017, in terms of log relative real hourly wages. Graph includes wage gap analysis for workers with high school relative to less than high school level, college or more relative to high school level and college or more relative to less than high school level. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

In terms of the evolution of the composition-adjusted experience premiums, Figure 7 provides an analysis of wage gaps between more experienced groups and new entrants (0-9 years of experience) from 2004 to 2017. Experience premium maintained a relative stable path, with some increase in 2011-2013 followed by a decrease in the last years. In particular, from 2015 to 2017, log relative real hourly wages of 20-29 years and 30 or more years groups reduced almost 0,04 log points (19%) and 0,04 log points (21%), respectively, after the increase in 2011-2013 of around 0,02 log points (14%) and 0,03 log points (20%), respectively. The significant reduction of log real hourly wages among the new entrants from 2011 to 2012 and the less significant increase from 2012 to 2013 (in comparison with the 20-29 years and 30 or more years groups), contributed for the observed increase in wage differentials between 2011 and 2013. After 2015, the more significant increase in wages of new entrants supported the decrease of wage differentials. Only log relative real hourly wages of 10-19 years group continuously diminished along the period, especially after 2013 where the observed reduction was around 0.04 log points (27%). The relatively stable wages of 10-19 years group and the increasing wages of new entrants after 2013 contributed for this pattern. Figure 7 also suggests that wage gaps between more experienced workers and new entrants contributed to the wage inequality pattern observed. The increase in wage gaps from 2011 to 2013 are in accordance with the slight increase in inequality during this period, contributing to the stability registered until 2013. From 2013 to 2017, the decrease in observed wage gaps matches the downward trend in inequality. Figure A2.3 in Appendix A2 replicates the analysis using ICOR datasets obtaining stable patterns with some fluctuations, until 2013, and a more consistent decrease after 2015.

Figure 7 – Evolution of composition-adjusted experience premium

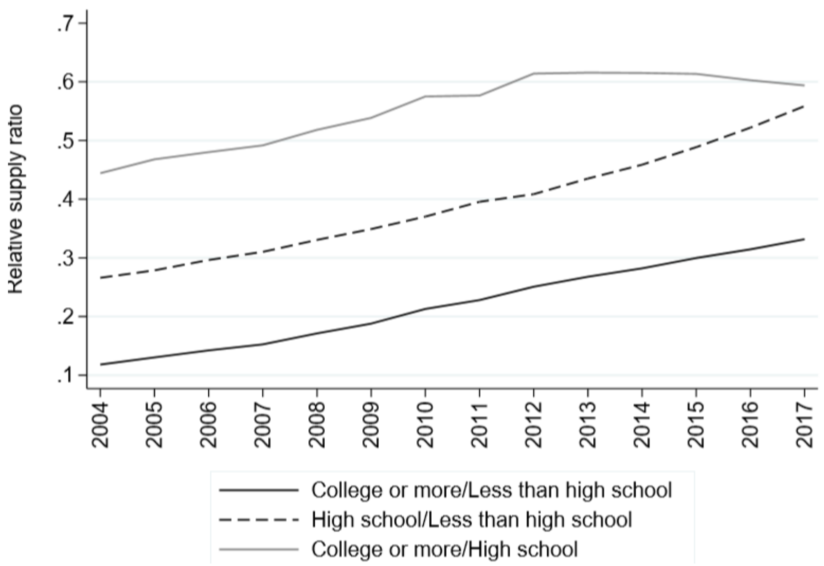


Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents composition-adjusted experience premium between 2004 and 2017, in terms of log relative real hourly wages. Graph includes wage gap analysis for workers with 10-19 years of experience relative to 0-9 years, 20-29 years of experience relative to 0-9 years and 30 or more years of experience relative to 0-9 years. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

The evolution of composition-adjusted education and experience premiums can be explained, among other factors such as minimum wage or unemployment, by the interaction of labor supply and demand for skills throughout the period. Recent studies, for example for the US, suggest that between the 1980's and 2000's, the lower growth of the supply of more educated workers played an important role on the increase of education premium (e.g. Goldin and Katz 2008). In Latin America, studies suggest that the increase in the relative supply of more educated and experienced workers in the 2000's is consistent with the decreasing wage differentials, even though demand side factors also played a crucial role (e.g. Messina and Silva 2019).

Figure 8 presents the relative labor supply between education groups in Portugal. The relative labor supplies are computed based on the relative number of workers in each education category and year. From Figure 8, the relative supply growth of more educated workers contributed to the reduction of the education premium registered between 2004 and 2017. In particular, the supply of college and high school educated workers grew relative to workers with less than high school level by 0,21 (almost triple) and 0,29 (more than double), respectively. These changes contributed to the decline in their premium. Although supply dynamics play a central role on the premium reduction between 2004 and 2017, demand for skills also contributes for the observed pattern. The supply ratio between college and high school educated workers increased almost 0,17 between 2004 and 2012, maintaining a relative stability in the following years. There is an important role of relative demand between college and high school educated workers to explain premium evolution, as the pattern observed on relative supply differs from that of wage differentials. Figure A2.4 in Appendix A2 presents the relative labor supply by education groups on ICOR datasets, obtaining similar increasing patterns for college and high school educated workers relative to less than high school educated workers supply.

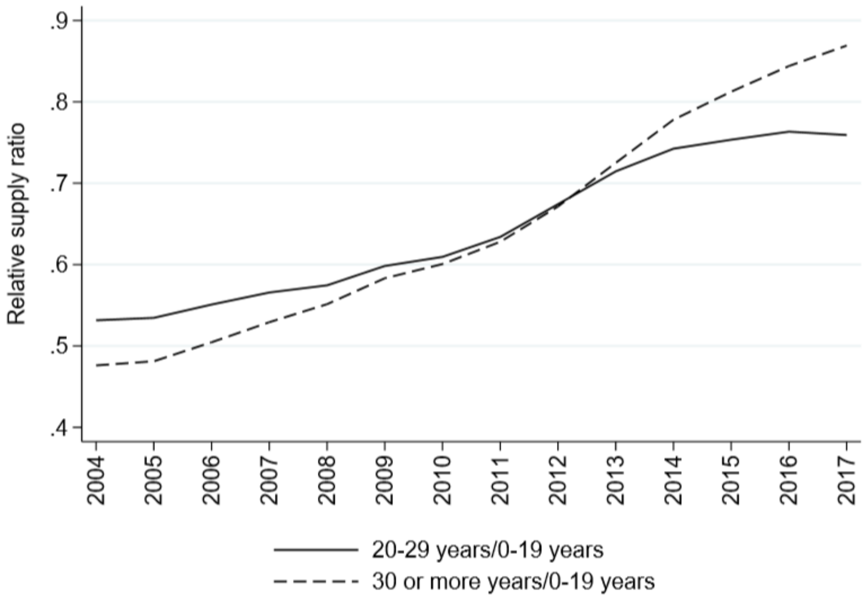
Figure 8 – Evolution of relative labor supply of different education levels



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents relative supply between education groups from 2004 to 2017. Graph includes relative supply analysis of workers with high school relative to less than high school level, college or more relative to high school level and college or more relative to less than high school level. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms.

Figure 9 presents the relative labor supply of different experience groups in Portugal, based on the relative number of workers by experience group. Between 2004 and 2017, the relative supply of workers with more years of experience grew significantly vis a vis that of the workers between 0-19 years of experience. The observed pattern indicates strong population aging effects (increasing supply of more experienced workers), surpassing the effect of a delayed labor market entry of younger cohorts due to an increasing number of schooling years. Given the increasing presence in the working age population of more experienced workers relative to less experienced, a significant decrease on experience premiums would be expected. However, the premiums only fell more significantly after 2015, emphasizing the important role played by the demand side factor for the stability of wage inequality in the initial years. Figure A2.5 in Appendix A2 replicates, for ICOR datasets, the analysis of relative labor supply by experience groups, indicating similar patterns.

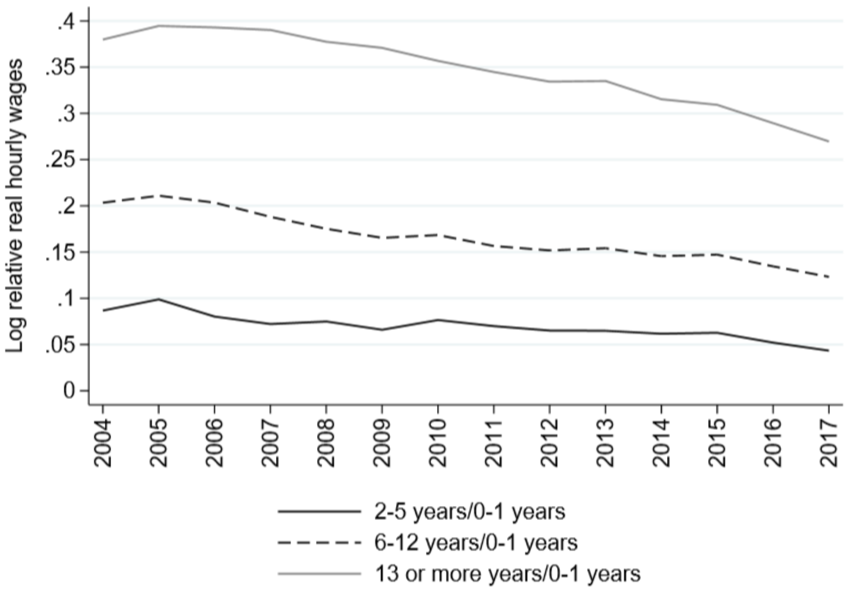
Figure 9 – Evolution of relative labor supply of different experience levels



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents relative supply between experience groups from 2004 to 2017. Graph includes a supply analysis of workers with 20-29 years relative to 0-19 years of experience and 30 or more years relative to 0-19 years of experience. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms.

Wage differentials between different education or experience groups are widely studied using age minus years of education to determine experience of workers (e.g. Autor, Katz and Kearney 2008; Rodríguez-Castelán et al. 2016). This measure indicates the overall experience and skills acquired throughout the professional career since entering the labor market. A complementary measure that I use is firm tenure, measured as the number of years working at the same firm. In the following analysis with firm tenure measuring experience, I still consider 24 education-experience-gender groups (skill groups), obtained from three education categories (less than high school education, high school education, college education or more), four experience categories in terms of tenure (0-1, 2-5, 6-12, 13 or more years of experience) and two genders[10]. The previous methodology to obtain composition adjusted experience premiums is repeated. Figure 10 presents the wage differentials between experienced workers and new entrants (0-1 years of experience) in terms of firm tenure between 2004 and 2017. The evolution of wage gaps suggests a reduction on earnings of more experienced workers relative to new entrants contributing to a reduction in wage inequality throughout the entire period. These patterns suggest the presence of countering effects before 2013 contributing to the stability of wage inequality registered.

Figure 10 – Evolution of composition-adjusted experience premium in terms of tenure

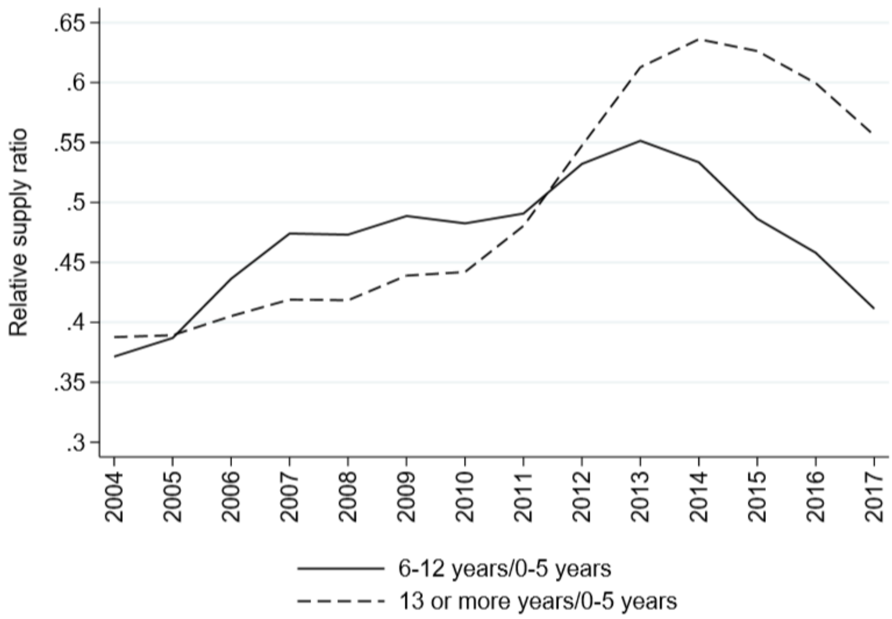


Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents composition-adjusted experience (tenure) premium between 2004 and 2017, in terms of log relative real hourly wages. Graph includes wage gap analysis for workers with 2-5 years of experience in the same current firm relative to 0-1 years, 6-12 years of experience in the same current firm relative to 0-1 years and 13 or more years of experience in the same current firm relative to 0-1 years. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

[10] Education is divided in three categories based on worker's highest education level completed. Worker's observations without information on education were eliminated. Workers that completed technical courses are considered as having completed high school education. Experience is given by the number of working years at the same current firm (tenure). The fixed weights used to compute the weighted average of predicted log of real hourly wages for broader education or experience groups are still the mean employment shares of each skill group between 2004 and 2017. Employment share of each group in year 2017 was also considered as an alternative weight, demonstrating similar results to the ones presented.

Figure 11 depicts the evolution of the relative labor supply between workers with different levels of firm tenure. The relative labor supply follows two distinct trajectories: before 2013/2014, the number of senior workers at firms increased relative to new entrants (0-5 years of experience at the firm) and, in the last years, this ratio has been falling. This changing pattern suggest an increasing trend in recent years for higher worker rotativity, where the number of new workers at firms increased significantly relative to the number of firm seniors. The increase in the relative labor supply of more experienced workers was fundamental for the reduction of composition-adjusted premiums before 2013. Nevertheless, the demand for new or senior workers at firm level may also have played an important role, explaining the premium decrease after 2013.

Figure 11 – Evolution of relative labor supply of different experience levels in terms of tenure



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents relative supply between experience (tenure) groups from 2004 to 2017. Graph includes a supply analysis of workers with 6-12 years relative to 0-5 years of experience in the same current firm and 13 or more years relative to 0-5 years of experience in the same current firm. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms.

This section also presents the results of a decomposition of the overall inequality in its between- and within-skill-groups components, quantifying the contribution of each component to changes on overall wage inequality.

Table 1 presents estimation results of equation 1 with the wage inequality decomposition. Columns 1, 3 and 5 present the total, between-skill-groups and within-skill-groups variance level with the respective share on total variance in columns 2, 4 and 6. Columns 7, 9 and 11 analyse the changes on total, between-skill-groups and within-skill-groups variances, having the percentage contribution to the change in columns 8, 10 and 12, respectively. Table 1 suggests that only around one third of the overall wage variance in each year occurs between skill groups. Within-skill-groups variance accounts for 72,7% of the observed reduction between 2009 and 2011, while between-skill-groups variance explains 56,7% of the continuous decrease and downward trend in wage inequality since 2013. Even though two thirds of total wage variance are explained by differences between workers with similar skills, differences across workers with distinct skills seems to play an important role for the latest reduction and downward trend on inequality. This pattern was not observed in similar studies for West Germany, where within-skill-groups variance played a more decisive role in inequality changes (e.g. Card, Heining and Kline 2013), or for Latin America, where the fall on inequality can be attributed to a major decrease in within-skill-groups variance (e.g. Messina and Silva 2019).

Table 1 – Variance of wages within and between skill groups

Year	Total log of real hourly wage variance		Between-skill-groups variance		Within-skill-groups variance	
	Variance component (1)	Share of total variance (2)	Variance component (3)	Share of total variance (4)	Variance component (5)	Share of total variance (6)
2004	0,2771	100	0,0908	32,8	0,1863	67,2
2005	0,2846	100	0,0947	33,3	0,1899	66,7
2006	0,2827	100	0,0940	33,2	0,1887	66,8
2007	0,2861	100	0,0941	32,9	0,1920	67,1
2008	0,2813	100	0,0945	33,6	0,1869	66,4
2009	0,2812	100	0,0963	34,2	0,1849	65,8
2010	0,2677	100	0,0926	34,6	0,1752	65,4
2011	0,2603	100	0,0906	34,8	0,1697	65,2
2012	0,2594	100	0,0928	35,8	0,1666	64,2
2013	0,2645	100	0,0941	35,6	0,1704	64,4
2014	0,2578	100	0,0899	34,9	0,1679	65,1
2015	0,2529	100	0,0856	33,8	0,1673	66,2
2016	0,2393	100	0,0787	32,9	0,1606	67,1
2017	0,2259	100	0,0722	32,0	0,1537	68,0

Year	Change on total log of real hourly wage variance		Change on between-skill-groups variance		Change on within-skill-groups variance	
	Variance component (7)	Share of total variance (8)	Variance component (9)	Share of total variance (10)	Variance component (11)	Share of total variance (12)
2004 - 2009	0,004	100	0,005	132,9	-0,001	-32,9
2009 - 2011	-0,021	100	-0,006	27,3	-0,015	72,7
2011 - 2013	0,004	100	0,004	84,4	0,001	15,6
2013 - 2017	-0,039	100	-0,022	56,7	-0,017	43,3

Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Tables present the decomposition of total variance of log real hourly wages in the variance between and within skill groups in each year. Besides providing a decomposition of variance, tables also provide the percentage contribution of between- and within-skill-groups variance to the overall variance. The focus continues to be on gross wage's inequality/variance among full-time dependent workers in the main occupation. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Workers with similar skills can receive different earnings as they work in different occupations and at different sectors. Table 2 presents the estimation results of equation 2 with the wage inequality decomposition[11]. Columns 1, 3, 5, 7, and 9 present the variance/covariance levels of each component, while columns 2, 4, 6, 8 and 10 present their respective share in total variance. Columns 11, 13, 15, 17, 19 display the changes on overall variance and on each of its components in levels, while their percentage contribution for the change on total variance is stated in columns 12, 14, 16, 18 and 20.

Table 2 suggests that, although around one fifth of the variance of log real hourly wages now occurs between sector-occupation groups, around one half continues to occur within skill and sector-occupation groups (residual component) and is not explained by differences across skills or sector-occupations. The positive covariance between skill and sector-occupation groups implies the existence of an association between highly paid skills, sectors and occupations. The residual component accounts for almost half of the observed reduction between 2010 and 2011 followed by the variance between sector-occupation groups explaining 27,6% of the decrease. Regarding the downward trend on overall wage variance from 2013 to 2017, variance between sector-occupations has a major importance, accounting for 30,4% of the decrease, followed by the residual component and the covariance between skill and sector-occupation groups each

[11] The worker's occupation and economic sector are classified according to the one-digit level classification of CPP/2010 and CAE Rev.3, respectively (more details on the classification systems and a list of the different one-digit sectors and occupations provided in Appendix A1.2). The analysis of the variance and its decomposition are only done from 2010 to 2017, due to the impossibility of harmonizing different classifications of occupations prior and after 2010 at the one-digit level, and it was eliminated worker's observations without job information.

explaining around 26%. From this table, it is possible to conclude that even when sector-occupation information is included, there is still a significant amount of inequality occurring within skill and sector-occupation groups. Nevertheless, most wage inequality evolution can be explained by differences across skill and sector-occupation groups. For example, there is a significant contribution for inequality reduction in the past years from the decrease in wage differences between sector-occupations and the association between highly remunerated skills, sectors and occupations.

Table 2 – Variance of wages within and between skill groups and sector-occupation groups

Year	Total log of real hourly wages variance		Variance in skill-groups		Variance in sector-occupation		2* Cov(skill-groups, sector-occupation)		Residual	
	Variance component (1)	Share of total variance (2)	Variance component (3)	Share of total variance (4)	Variance component (5)	Share of total variance (6)	Covariance component (7)	Share of total variance (8)	Variance component (9)	Share of total variance (10)
2010	0,2677	100	0,0325	12,2	0,0568	21,2	0,0419	15,7	0,1365	51,0
2011	0,2603	100	0,0316	12,1	0,0547	21,0	0,0412	15,8	0,1328	51,0
2012	0,2594	100	0,0318	12,3	0,0553	21,3	0,0423	16,3	0,1299	50,1
2013	0,2644	100	0,0317	12,0	0,0580	21,9	0,0427	16,2	0,1320	49,9
2014	0,2578	100	0,0307	11,9	0,0558	21,7	0,0406	15,7	0,1306	50,7
2015	0,2528	100	0,0295	11,7	0,0536	21,2	0,0383	15,1	0,1314	52,0
2016	0,2392	100	0,0273	11,4	0,0500	20,9	0,0353	14,7	0,1266	52,9
2017	0,2258	100	0,0252	11,2	0,0463	20,5	0,0325	14,4	0,1219	54,0

Year	Change on total log of real hourly wage variance		Change on variance in skill-groups		Change on variance in sector-occupation		Change on 2*Cov(skill-groups, sector-occupation)		Change on residual component	
	Variance component (11)	Share of total variance (12)	Variance component (13)	Share of total variance (14)	Variance component (15)	Share of total variance (16)	Covariance component (17)	Share of total variance (18)	Variance component (19)	Share of total variance (20)
2010 - 2011	-0,0074	100	-0,0009	12,6	-0,0021	27,6	-0,0007	10,0	-0,0037	49,8
2011 - 2013	0,0041	100	0,0001	1,9	0,0033	79,4	0,0016	38,2	-0,0008	-19,4
2013 - 2017	-0,0386	100	-0,0065	16,8	-0,0117	30,4	-0,0102	26,6	-0,0101	26,3

Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Tables present the decomposition of total variance of log real hourly wages, in each year, in the variance in skill groups and sector-occupation groups, twice the covariance between the two previous terms and the variance of residual component (referred simply in the table as residual). Besides providing a decomposition of variance, tables also provide the percentage contribution of each component to the overall variance. The focus continues to be on gross wage's inequality/variance among full-time dependent workers in the main occupation. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

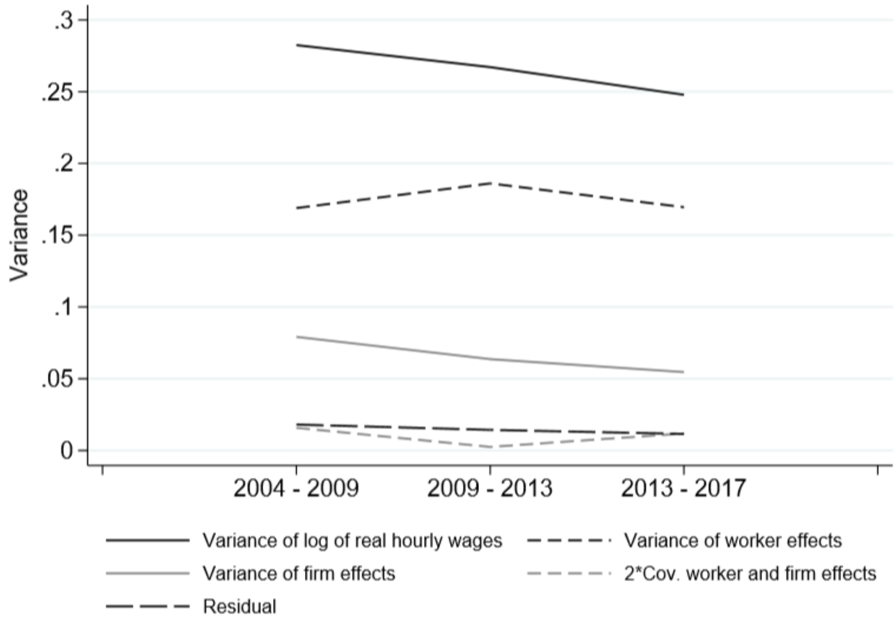
4.3. Heterogeneity across workers and firms

The decomposition of total wage variances in each subperiod (2004-2009, 2009-2013 and 2013-2017) (equation 4) are presented in Figure 12 and with more detailed information on Table A2.3 in Appendix A2. Figure 12 shows that the variance of worker effects, associated with heterogeneity across workers, accounts for the majority of the variance of log real hourly wages in the three subperiods (between 60% and 70% of total variance), followed by firm effects that explain 22% to 28% of overall wage variance.

Regarding the variation of total wage variance and its components, two patterns emerge. First, in the two initial subperiods from 2004 to 2013, the combination of a decrease in the variance of firm effects and in the doubling of the covariance between worker and firm effects contributed to around 101% and 87% of the reduction on overall wage variance (mainly between 2009 and 2011), respectively. However, the observed decrease in inequality was smaller due to the countering effect of increased variance of worker effects. Thus, from 2004 to 2013, the reduction on heterogeneity across firm's pay premiums to their workers and on the association between highly remunerated workers and high-paying firms surpassed the increase on heterogeneity across workers, leading to a decrease in wage inequality. Nevertheless, higher variance of worker fixed effects (heterogeneity across workers) contributed to a more stable trend of wage inequality until 2013, with only a mild reduction between 2009 and 2011, countering, for example, the effects of the reduction of education or experience (tenure) premiums documented above.

Second, in the last two subperiods from 2009 to 2017, there was a reduction in the variances of worker and firm effects that contributed to 86% and 47% of the observed decrease on total wage variance (especially after 2013 when the reduction of inequality was more pronounced), respectively. However, the opposite effect of the increase in the covariance between worker and firm effects led to a lower reduction in the overall variance. Thus, from 2009 to 2017, the reduction on heterogeneity across workers (from wage differences between fixed worker characteristics that are paid similarly across firms) and firms (from wage differences between employers related to firm's pay premiums to their employees) surpassed the effect of increasing association between highly remunerated workers and high-paying firms. This allows a reduction of wage inequality and contributes to the downward trend verified after 2013. From the analysis of the entire period, the decreasing variance of firm effects was the main responsible for the observed decrease on overall variance between 2004 and 2017, accounting for almost 71% of the variation. The variance of residual also accounted for around 19% of the reduction on overall variance between 2004 and 2017, followed by the covariance between worker and firm effects accounting for 11% of the change.

Figure 12 – Decomposition of variance of wages across workers and firms



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents the decomposition of total variance of log real hourly wages, in each subperiod, in the variance of worker and firm effects, twice the covariance of worker and firm effects and variance of the residual (referred simply in the graph as residual). The variance of year effects, twice the covariance of worker and year effects and twice the covariance of firm and year effects are also part of this decomposition, but they were omitted from the graph due to their insignificant role. The focus continues to be on gross wage's inequality/variance among full-time dependent workers in the main occupation. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed. More details on the decomposition are provided in Table A2.3 in Appendix A2.

The results obtained for Portugal hold some different patterns when compared with recent studies done for other countries that also decomposed variance using AKM models. For example, recent studies in Brazil (e.g. Alvarez et al. 2018) indicate that the variance of worker effects is the main component (as in Portugal) accounting for 58% of overall variance between 2008 and 2012, followed by the doubling of covariance of worker and firm effects and the variance of firm effects accounting 18% and 16%, respectively, in the same period. The majority of the decrease in wage inequality (around 40%) from 1996 to 2012 is due to the firm component, as in Portugal. However, unlike the Portuguese pattern, worker component and covariance of worker and firm effects have a high weight of 29% and 23% in the reduction of wage inequality, respectively. Regarding the US example, recent studies (e.g. Song et al.

2019[12]) suggest that the variance of worker effects is the largest component, as in Portugal, explaining around 52% of wage variance between 2007 and 2013. Nevertheless, the firm component explains only 9% of the total variance (significantly less than in Portugal), while the covariance of worker and firm effects explains around 12% (significantly more than in Portugal). In terms of the variation of wage variance between 1980 and 2013, differently from Portugal, the variance of worker effects accounted for almost 68% of the rise in overall variance followed by the covariance of worker and firm effects contributing for 35% of the increase. The variance of firm effects slightly decreases throughout the period as in Portugal, contradicting the increase in wage inequality.

The previous analysis of wage inequality using AKM models to decompose variance in a worker and firm component, besides other components, does not provide a direct measure on inequality occurring within firms. If each firm promotes a more homogeneous wage policy with less differences across their employees, wage inequality will decrease even if the heterogeneity across firms is maintained. Thus, the second part of this section will analyse specifically the role of heterogeneity within firms for the variation in wage inequality, by decomposing total variance on the between- and within-firm's components.

Before decomposing the overall variance of log real hourly wages in the between- and within-firm components in each year, wages are separated, following Alvarez et al. (2018); Song et al. (2019) and Messina and Silva (2019) approach, in:

$$wage_{ift} \equiv \overline{wage_t} + \left(\overline{wage_{ft}} - \overline{wage_t}\right) + \left(wage_{ift} - \overline{wage_{ft}}\right) \tag{5}$$

where $wage_{ift}$ is the log of real hourly wages of worker i in firm f at year t, $\overline{wage_t}$ is the average log of real hourly wage in the economy at year t, $\overline{wage_{ft}}$ is the average log of real hourly wage in firm f at year t. The wages of each worker can be interpreted as a result of the average remuneration on the economy in a certain year plus the difference paid on average by firms relative to the average on the economy plus the difference earned by workers relative to the average of their firms. In order to obtain the within- and between-firm components of wage variance in each year, identity (5) is transformed, using variances, in:

[12] This paper analyses variance decomposition in US for men observations, while the variance in Portugal is analysed for men and women observations together.

$$Var\left(wage_{ift} - \overline{wage_t}\right) = Var\left(\overline{wage_{ft}} - \overline{wage_t}\right) + Var\left(wage_{ift} - \overline{wage_{ft}}\right) \quad (6)$$

$$\text{with } 2 * Cov\left(\overline{wage_{ft}} - \overline{wage_t};\ wage_{ift} - \overline{wage_{ft}}\right) = 0$$

Taking into account that variance is decomposed for each year, $\overline{wage_t}$ becomes a constant in equation (6) and it can be simplified in:

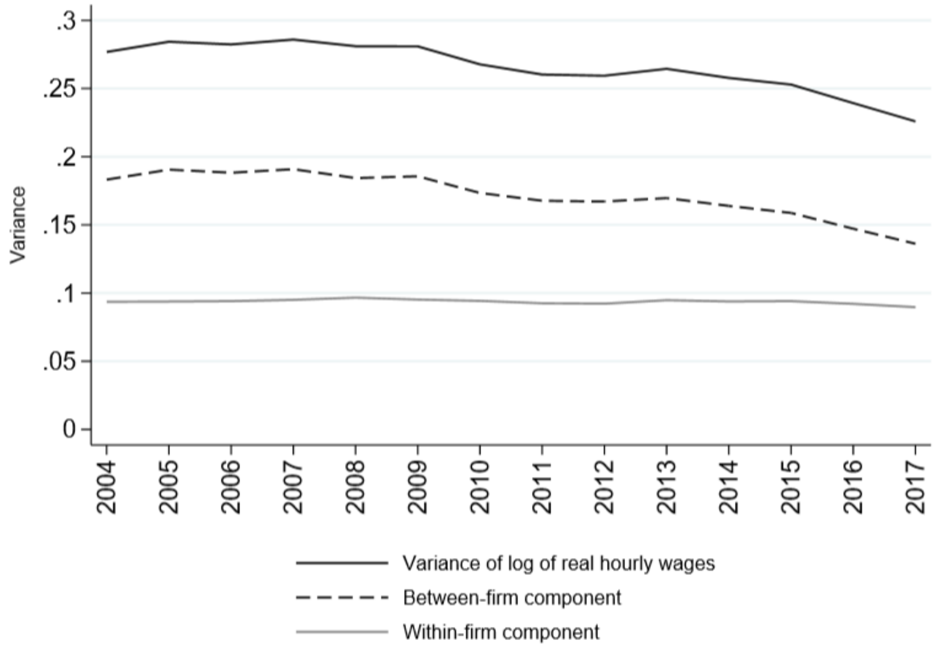
$$Var\left(wage_{ift}\right) = Var\left(\overline{wage_{ft}}\right) + \overline{Var\left(wage_{\iota ft} \mid \iota \in f\right)} \quad (7)$$

Equation (7) decomposes, in each year, the overall variance of log real hourly wages in the between-firm component given by the variance of firm's average wages and the within-firm component given by the average variance of workers' wages within each firm.

After decomposing wage inequality using equation (7), the variance of wages between and within firms are estimated for each year and the results are presented in Figure 13. From the analysis of Figure 13, two main conclusions emerge. First, the majority of total wage variance comes from the between-firm component in every year, indicating that wage differences across firms have a higher weight in wage inequality than the differences within firms. In particular, the variance between firms account for more than 60% of total variance in each year.

Second, the within-firm component of the variance of log real hourly wages was relatively stable throughout the period, contributing less for the evolution of wage inequality. Thus, around 92% of the wage inequality variation on the period was due to changes in dispersion across firms. In particular, the variance between firms accounted for 87% of both the decrease of inequality from 2009 to 2011 and the reduction after 2013, indicating small improvements in terms of inequality between workers of the same firm. Recent studies for Brazil indicate similar results, with between-firm wage variance being the main responsible for the observed fall in wage inequality since 1996 (e.g. Alvarez et al. 2018; Messina and Silva 2019). In the United States, studies also point out that, although within-firm wage variance is the largest component of total variance, the between-firm component contributed more to the increasing pattern of wage inequality between 1978 and 2013 (e.g. Song et al. 2019).

Figure 13 –Variance of wages between and within firms



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents the decomposition of total variance of log real hourly wages, in each year, in between-firm variance and within-firm variance. The focus continues to be on gross wage's inequality/variance among full-time dependent workers in the main occupation. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

5. Wage inequality and business cycle

In previous sections, the evolution of wage inequality in Portugal from 2004 to 2017 and the potential drivers of the observed pattern related to workers and firms were studied. Since 2004, Portugal faced crisis and non-crisis periods that affected worker's wages and wage inequality. Section 5 studies the wage inequality patterns along the business cycle, through the changes of wages during different subperiods. The main conclusion obtained was that wage reduction in higher and lower percentiles during the crisis subperiod contributed for a stable inequality trend. The wage increases at lower percentiles and decreases at higher percentiles during the subperiod of positive economic growth contributed to a downward trend on inequality.

Following Autor, Katz and Kearney (2008); Centeno and Novo (2014) and Messina and Silva (2019) approach when studying polarization in the United States, Portugal and Latin America, respectively, variations in wages were measured by sector-occupation wage percentiles from 2010 to 2013 and 2014 to 2017. Sector-occupation groups were formed using worker's occupation and economic sector classified according to the one-digit level of CPP/2010 and

CAE Rev.3, respectively. Since the classification system used for occupations changed from 2009 to 2010 and it is not possible to harmonize the two classifications at the one-digit level, the analysis was done only from 2010 onwards. The wage percentiles were obtained by ranking average wages of sector-occupation groups in the base year of each interval. Sector-occupation groups that have no information on worker's job or have less than 20 observations in a certain year of the interval were dropped from the analysis. Since Portugal faced a recession from 2009 to 2013 and an expansion between 2014 and 2017 (based on real GDP growth[13]), wage variations are studied in these intervals (except for the first interval where analysis only starts at 2010). The wage variation by sector-occupation wage percentile, in each interval, was obtained by the difference between the average log of real hourly wages of each sector-occupation group in the last year of the interval and the average log of real hourly wages of the same sector-occupation group in the initial year of the interval. After computing the changes on log of real hourly wages, it was determined the smoothed values of the changes in wages using a locally weighted smoothing regression with a default bandwidth of 0,8.

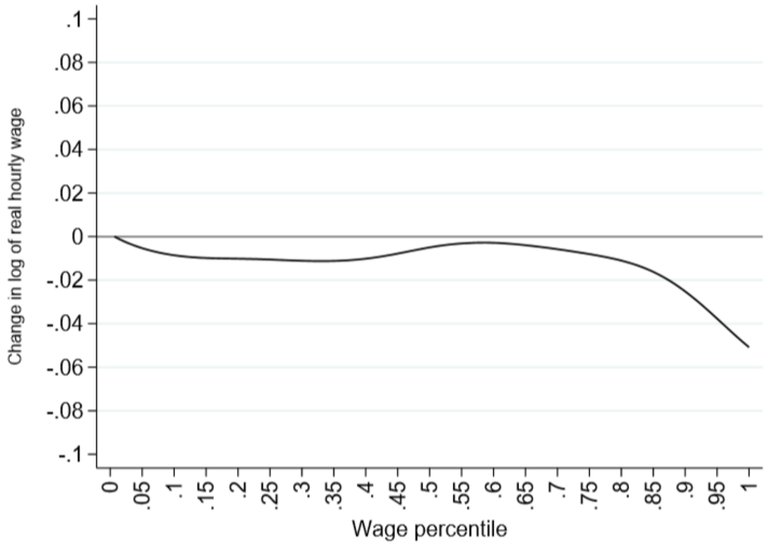
The results for the changes on log of real hourly wages between 2010 and 2013 and from 2014 to 2017 by sector-occupation wage percentile, are presented in Figure 14. From the analysis of wage variations during the recession and expansion periods, wage inequality trends can be explained. During the recession period from 2010 to 2013, a reduction on real hourly wages was verified, in accordance with the stylized facts on the procyclicality of real wages. In Figure 14, from 2010 to 2013, the observed decrease of wages occurred mainly at top percentiles, but also with some expression at lower percentiles. In particular, above the 75th percentile there was an increasing reduction of wages, reaching a 0,04 log points decrease at the 95th percentile, while below the 50th percentile there was a decrease of around 0,01 log points. Thus, the pattern of wage variation during the recession period, where wages decreased at lower and higher wage percentiles, supports the relative stability in wage inequality until 2013 (despites slight decrease in 2010 and 2011). During the expansion period from 2014 to 2017, an increase on real hourly wages at bottom percentiles, consistent with the procyclicality of real wages, and a decrease at top percentiles were witnessed. In particular, below the 60th percentile, there was a higher increase of wages at bottom percentiles, achieving an increase of around 0,06 and 0,04 log points at the 5th percentile and 20th percentile, respectively. Above the 60th percentile there was an opposite pattern, with a higher reduction of wages at top percentiles, reaching 0,06 and 0,04 log points decrease at the 95th percentile and 90th percentile, respectively. Thus, the combined

[13] Real GDP growth rates data obtained from PORDATA and computed by INE, PORDATA

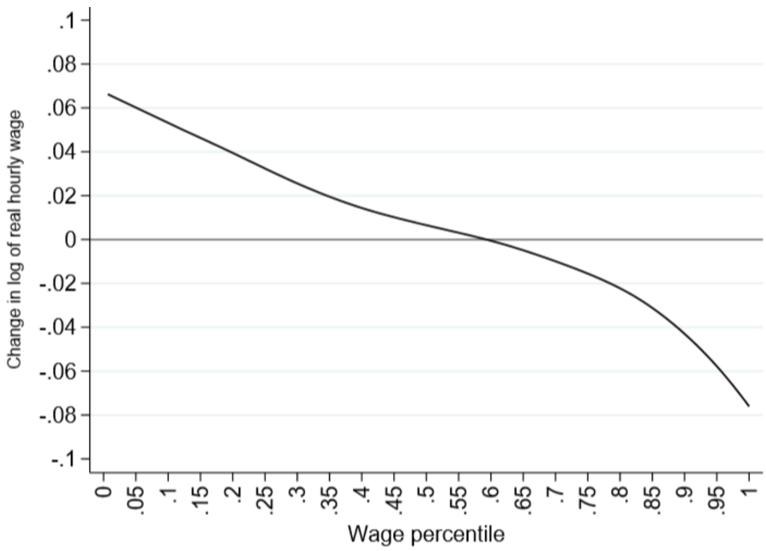
effect of an increase of wages in sector-occupations earning on average less and a decrease of wages in sector-occupations earning on average more, contributed for the downward wage inequality trend after 2013. Regarding the influence of business cycles in wage inequality, it can be concluded that inequality maintained a stable pattern during the last recession in Portugal due to a decrease at lower and higher sector-occupation wage percentiles. During the current expansion period, inequality fell significantly due to an increase of wages in poorer sector-occupations and decrease at richer ones.

Figure 14 – Wage inequality in crisis and non-crisis subperiods

2010 - 2013



2014 - 2017



Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents the smoothed values for the changes in log real hourly wages by sector-occupation wage percentile. The wage percentiles were obtained by ranking sector-occupation groups according to the average log of real hourly wages in the initial year of each interval. The wages used are full-time dependent workers' wages in gross terms in the main occupation. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

6. Conclusion

Using Portuguese harmonized household survey data (ICOR) and linked employer-employee datasets (QP), this paper studies the evolution of wage inequality in Portugal between 2004 and 2017 (2018 when using ICOR). Until 2013, inequality was relatively stable, with an observed decrease between 2009 and 2011 and a slight increase in 2012 and 2013. After 2013, wage inequality decreased continuously, contributing to the existence of a downward trend. Nine main conclusions regarding wage inequality patterns and potential drivers were documented.

First, the faster growth in bottom percentiles relative to top percentiles led to a gradual reduction of wage inequality after 2013. Second, Portugal was one of the European countries in which inequality fluctuated the most and diminished, similarly to Central and Eastern European countries. Third, in terms of age groups, inequality fell since 2004 among younger groups, while for older groups it only decreased after 2013 (when wage inequality pattern started to be downwards). Fourth, education premiums and experience premiums related to tenure decreased since 2004, contributing to lower inequality, while experience premiums were stable and only decreased after 2015. Fifth, even though within-skill-groups variance accounts for two thirds of wage variance, the between-skill-groups variance was fundamental to reduce wage variance after 2013. Sixth, even if variance between sector-occupations was important to reduce wage inequality after 2013, the variance within skills and sector-occupations is still the major component on total variance and with a large weight on the inequality pattern. Seventh, using AKM models to decompose wage variance, firm effects account for the majority of the decrease in wage variance throughout the period. Worker effects increase between 2004 and 2013 was the main responsible for a stability trend in inequality, contradicting the decrease on premiums and heterogeneity across firms. Eighth, the between-firm wage inequality accounts for the majority of the wage inequality level and evolution in comparison with the within-firm wage dispersion. Ninth, the wage decreases along different wage percentiles during a crisis subperiod promoted inequality stability, while wage increases at lower percentiles and decreases at higher percentiles contributed to a downward trend in inequality during an expansion period.

Thus, this paper provides new evidence on the potential drivers of the wage inequality changes in Portugal, including the role of from worker's observable characteristics, sector-occupation characteristics and worker and firm heterogeneity. However, the role of institutional factors as the minimum wages were not debated. Recent increases in the latter may contribute to lower wage inequality. Also, worker's characteristics, that are not associated to their skills or occupation, such as the gender or the geographical area and their gaps were not studied directly, even though worker fixed effects might have accounted for them when decomposing wage inequality. Regarding firms, this paper also does not provide a direct analysis of the heterogeneity across and within smaller/larger firms (as Song et al. 2019 done for US) or across and within better/worse performing firms, analysing all employers without division in cohorts. These topics constitute an avenue for further research on the drivers of wage inequality in Portugal, in particular their contribution for the recent downward trend. In terms of the analysis of wage inequality and business cycles this thesis offered brief study of inequality dynamics in crisis and non-crisis. This is an area with potential for more research quantifying and estimating how inequality changes with business cycles. Finally, the reason for the significant wage inequality decrease, since 2004, among the new entrants and younger age groups, was not investigated. This phenomenon might be related to lower education premiums or heterogeneity across firms and is an important avenue for future research.

References

Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Technical Paper 2002-06. Center for Economic Studies, US Census Bureau.

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2), 251-333.

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter and D. Card, editors, *Handbook of labor economics* (Vol. 4, Part B, pp. 1043-1171). Elsevier.

Alvarez, J., Benguria, F., Engbom, N., & Moser, C. (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics*, 10(1), 149-89.

Alves, N., Cardoso, F., & Monteiro, N. (2020). Uma caracterização da desigualdade do rendimento e do consumo em Portugal. *Banco de Portugal, Revista de Estudos Económicos*, 6(1), 4-24.

Arnold, J. M., & Farinha Rodrigues, C. (2015). Reducing inequality and poverty in Portugal. OECD Economics Department Working Papers, No. 1258. *SSRN Electronic Journal*

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory*, 2(3), 244-263.

Atkinson, A. B. (2008). More on the measurement of inequality. *Journal of Economic Inequality*, 6(3), 277-283.

Autor, D. H., & Dorn, D. (2013). How Technology Wrecks the Middle Class. *The New York Times*, August 24

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6), 2121-2168.

Autor, D. H., Dorn, D., Hanson, G. H., & Song, J. (2014). Trade adjustment: Worker-level evidence. *The Quarterly Journal of Economics*, 129(4), 1799-1860.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of economics and statistics*, 90(2), 300-323.

Autor, D. H., Katz, L. F., & Krueger, A. B. (1998). Computing inequality: Have computers changed the labor market?. *The Quarterly journal of economics*, 113(4), 1169-1213.

Bachmann, R., Bechara, P., & Schaffner, S. (2016). Wage Inequality and Wage Mobility in Europe. *Review of Income and Wealth*, 62(1), 181-197.

Brandolini, A., Rosolia, A., & Torrini, R. (2011). The distribution of employees' labour earnings in the European Union: Data, concepts and first results. *ECINEQ WP*, 198.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1), S13-S70.

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly journal of economics*, 128(3), 967-1015.

Cardoso, A. R., Guimaraes, P., Portugal, P., & Reis, H. (2018). The returns to schooling unveiled. *IZA Discussion Paper No. 11419*.

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

Correia, S., (2014). REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components S457874, Boston College Department of Economics, revised 18 Nov 2019.

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: A bad mix. *Technical Note, Duke University*.

Farinha Rodrigues, C., & Andrade, I. (2013). Growing Inequalities and Their Impacts in Portugal. *SSRN Electronic Journal*.

Feenstra, R. C., & Hanson, G. H. (1999). The impact of outsourcing and high-technology capital on wages: estimates for the United States, 1979–1990. *The Quarterly Journal of Economics*, 114(3), 907-940.

Ge, S., & Yang, D. T. (2014). Changes in China's wage structure. *Journal of the European Economic Association,* 12(2), 300-336.

Gerard, F., Lagos, L., Severnini, E., & Card, D. (2018). Assortative matching or exclusionary hiring? The impact of firm policies on racial wage differences in Brazil. *NBER Working Paper No. 25176*.

Goldin, C. D., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, United States: The Belknap Press of Harvard University Press.

Guimarães, P., & Portugal, P. (2010). A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects. *Stata Journal*, 10(4), 628-649.

Instituto Nacional de Estatística. (2004 to 2018). Inquérito às Condições de Vida e Rendimento.

Instituto Nacional de Estatística. (2007). *Classificação Portuguesa das Actividades Económicas Rev.3*. Lisbon.

Instituto Nacional de Estatística & PORDATA. (2010 to 2017). Real GDP growth rate. Accessed on 10th March 2020 on: https://www.pordata.pt/en/Portugal/Real+GDP+growth+rate-2298

Instituto Nacional de Estatística. (2011). *Classificação Portuguesa das Profissões 2010*. Lisbon.

Instituto Nacional de Estatística & PORDATA. (2017). Gross disposable income of households. Accessed on 10th March 2020 on: https://www.pordata.pt/en/Portugal/Gross+disposable+income+of+households+(2016)-2407

Instituto Nacional de Estatística. (2019). Documento Metodológico. 102 - Inquérito às Condições de Vida e Rendimento, version 3.7. Accessed on 4th December 2019 on: http://smi.ine.pt/DocumentacaoMetodologica/Detalhes/?id=1547&lang=PT

Kumler, T., Verhoogen, E., & Frías, J. A. (2013). Enlisting employees in improving payroll-tax compliance: Evidence from Mexico. *NBER Working Paper No. 19385*.

Krueger, A. (2012). The rise and consequences of inequality. Presentation made to the Center for American Progress, Washington, DC, United States, January 12.

Lee, J. W., & Wie, D. (2015). Technological change, skill demand, and wage inequality: Evidence from Indonesia. *World Development*, 67, 238-250.

Lee, J. W., & Wie, D. (2017). Wage structure and gender earnings differentials in China and India. *World Development*, 97, 313-329.

Magda, I., Gromadzki, J., & Moriconi, S. (2019). Firms and wage inequality in Central and Eastern Europe. *IZA Discussion Paper No. 12214*.

Messina, J., & Silva, J. C. G. (2019). Twenty Years of Wage Inequality in Latin America. Policy Research Working Paper 8995. Washington, DC, United States: World Bank.

Ministério do Trabalho e da Solidariedade. (2005). Documento Metodológico. 395 – Quadros de Pessoal, versão 1. Accessed on 4th December 2019 on: http://smi.ine.pt/DocumentacaoMetodologica/Detalhes/771

Ministério do Trabalho, da Solidariedade e da Segurança Social. (2004 to 2017). Quadros de Pessoal.

Ministério do Trabalho, da Solidariedade e da Segurança Social. (2016). Quadros de Pessoal (Relatório Único – Anexo A). *Estatísticas em Síntese*. Accessed on 5th January 2020 on: http://www.gep.mtsss.gov.pt/documents/10182/10928/qp2016sint.pdf/1ef72e73-4f48-43a3-b1b3-863bb5c51eaa

Mittag, N. (2016). A Simple Method to Estimate Large Fixed Effects Models Applied to Wage Determinants and Matching. *IZA Discussion Paper No. 10447.*

OECD.Stat. (2020). Decile ratios of gross earnings. Accessed on 23th February 2020 on: https://stats.oecd.org/Index.aspx?DataSetCode=DEC_I

OECD. (2020). Inflation (CPI) (indicator). doi: 10.1787/eee82e6e-en (Accessed on 23 February 2020).

Pierce, J. R., & Schott, P. K. (2016). The surprisingly swift decline of US manufacturing employment. *American Economic Review*, 106(7), 1632-1662.

Portugal, P., Raposo, P. S., & Reis, H. (2018). Sobre a distribuição de salários e a dispersão salarial. *Banco de Portugal. Revista de Estudos Económicos*, 4(1), 1-14.

Regulation (EC) No 1177/2003 of June 16th 2003. *Official Journal of the European Union, L 165, July 3rd 2003, pp. 1-9*. European Parliament and Council.

Robin, J. M. (2011). On the dynamics of unemployment and wage distributions. *Econometrica*, 79(5), 1327-1355.

Rodríguez-Castelán, C., López-Calva, L. F., Lustig, N., & Valderrama, D. (2016). Understanding the dynamics of labor income inequality in Latin America. Policy Research Working Paper 7795. Washington, DC, United States: World Bank.

Song, J., Price, D. J., Guvenen, F., Bloom, N., & Von Wachter, T. (2019). Firming up inequality. *The Quarterly journal of economics*, 134(1), 1-50.

Torres, S. M., Portugal, P., Addison, J. T., & Guimarães, P. (2013). The sources of wage variation: a three-way high-dimensional fixed effects regression model. *IZA Discussion Paper No. 7276.*

Appendix

A1 – Data Sources

A1.1 Inquérito às Condições de Vida e Rendimento

During part of Section 4.1 as well as in part of the following Appendix A2, annual data since 2004 from one Portuguese household survey called *Inquérito às Condições de Vida e Rendimento* (ICOR) is used. The Portuguese Institute of Statistics (INE) is the entity responsible for this operation together with DES/CV (organic units) and with external relation to EUROSTAT and Unit F-4: Quality of life (another organic unit).

The survey and database arose from the need, at the European Union level, to obtain statistics and indicators on income and living conditions, comparable between the different member-states, to help decision-making in the field of social cohesion by European leaders. Thus, ICOR is due to the implementation in Portugal of the European community statistical database called Statistics on Income and Living Conditions (EU-SILC) according to the European Parliament and Council Regulation (EC) Nº 1177/2003 of 16th June 2003. Among other things, this database allows to study "the composition and distribution of the income of families and individuals" (INE, *Inquérito às Condições de Vida e Rendimento*, Methodological Document, 2019, p.10).

The target of the survey are all the people living in national territory (including Madeira e Azores) and a sample of the population is collected using housing as sample units. The sample of the population is carefully drawn (and not simply random) with "stratification and selection of units in different steps: areas first and housing afterwards" (INE, *Inquérito às Condições de Vida e Rendimento*, Methodological Document, 2019, p.16). *Ficheiro Nacional de Alojamentos* (FNA) is used in order to obtain the sampling basis that is used afterwards to select the annual sample of ICOR. In this sample, the statistical unit of observation is not only the household but also the individual. There are four subsamples composing the sample in each year and in each year one subsample is replaced by another on the same area, in such a way that a subsample does not remain more than four consecutive years in this survey. This sampling model allows not only to obtain cross-sectional data on households and individuals, but also longitudinal data as it follows the households and individuals on a subsample during 4 consecutive years.

After building the questionnaire of the survey, there are face-to-face interviews between March and June with the selected households and individuals and collection of their income information on the previous year. Afterwards, the information is analysed and validated to ensure the structure and the coherency of the data and it is also anonymized and attributed multiple weights to the household and individual (for the individual there is a basis weight, a weight for cross-sectional analysis and yet other for longitudinal analysis). Since the beginning of ICOR in 2004, there are regular updates and reviews of the questionnaires of the survey, made by the responsible entities or related to changes on technical guidelines at the European Union level. There were also changes related to the size, rotation of the sample collected and classification system of some variables like education and occupations, as well as faster and more efficient procedures after collecting the data, allowing sooner dissemination of the results of the survey.

In this thesis, cross-sectional individual data from ICOR between 2004 and 2018 (last year with the results available) is used. The data of the different years was first appended and treated to ensure that the same variables in different years have equal denomination. Since the target of the study were full-time dependent workers aged between 18 and 65 years old (working age population), only individual observations reporting the following were selected: age between 18 and 65 years old, self-definition of working conditions corresponding to full-time dependent workers (although, before 2009, this variable does not separate full-time dependent and independent workers), job situation corresponding to dependent worker and main activity in each month corresponding to full-time dependent workers. Since the purpose is to study wage inequality on full-time dependent workers, the variable of gross monthly earnings (in main activity) from dependent work in the year of the interview, which includes the "usual paid overtime, tips, commissions and a proportionate share of supplementary payments like the 13th month payment or an annual bonus" (Brandolini, Rosolia, and Torrini 2011, p. 4) was used. After eliminating the observations with zero or missing values in this variable, the gross monthly earnings from dependent work were divided by four and converted in gross weekly labor earnings (assuming 4 weeks per month). Gross hourly labor earnings were obtained by dividing the gross weekly earnings by the number of weekly working hours. To convert nominal hourly earnings in real hourly earnings, the consumer price index (CPI)[14] deflators in the different years for Portugal were used.

[14] The CPI values were not part of ICOR and were obtained from OECD.

In terms of highest completed or currently enrolled education level, the system of classification used is the International Standard Classification of Education (ISCED). In particular, before 2014 is used ISCED-97 and in 2014 or after is used the ISCED-11. Although there are changes in the version of ISCED, including at the one-digit level classifications, it is still possible the aggregation of different groups at different years in individuals with: i) less than high school education, ii) high school education level and iii) college education or more, based on the variable of education level completed or enrolled (when there are missing values on the completed education). The experience of the workers is given by the number of years of paid work variable.

Besides treating the datasets of different years, the 1st and 99th percentile of real hourly wages in each year were trimmed to eliminate outliers on the wage distribution.

A1.2 Quadros de Pessoal

For the majority of the analysis done, including from part of Section 4.1 to Section 5, the annual data since 2004 from Portuguese administrative linked employer-employee datasets called *Quadros de Pessoal* (QP) is used. The entity responsible for this statistical operation with national geographical scope is *Gabinete de Estratégia e Planeamento* (GEP) from the Ministry of Employment, Solidarity and Social Security (MTSSS), making the data available to INE.

The panel data needed for this dataset is obtained through an annual administrative census, where employers with at least one dependent worker are required to deliver, electronically or manually, to the responsible entity the information on the employees and their earnings (for example gender, highest education level completed, job titles, collective bargaining agreement, date of birth, occupation, date of hiring), as well as, information on the firms (for example, equity, number of employees, sales, sector of activity) and establishments, as a way to verify if they are complying with labor law. The only employers that are not required to deliver information, besides the self-employed individuals, are "central, regional and local administration and public institutes (for these entities only applicable to workers under individual employment contract) and to employers of domestic service workers" (MTSSS, *Quadros de Pessoal*, Methodological Document, 2005, p.1), covering virtually all firms and their dependent workers in Portuguese private sector. All data is referent to October and should be delivered in November every year (annual data). The units of observation might be the worker, firm or establishments depending on the information. However, datasets of different years and units of observation are relatable through the use of unique identifiers for workers, firms and establishments, that allow to follow them through time (longitudinal data) and associate each worker with an establishment and a firm (linked employer-employee datasets).

This paper uses worker longitudinal data from QP between 2004 and 2017 (last year with data available). The data on each year was first treated by merging firm and worker datasets, using their unique identifiers, and it was verified the classification system of variables like sector or occupation. Worker's observations having a worker ID with less than 6 digits or more than 10 digits are invalid and were discarded. When there were duplicated observations on a certain year for a certain worker ID, only the observation with highest normal monthly hours of paid work and highest gross monthly basic earnings (more likely to be the worker's main job) was kept. Since the purpose is to study wage inequality on full-time dependent workers aged between 18 and 65 years old (working age population), it was only kept, at each year, observations of workers with: an age between the 18 and 65 years old, job situation corresponding to dependent worker and at least 120 normal monthly hours of paid work. At each year, it was eliminated observations of workers without complete basic remuneration, belonging to residual categories on job titles or that belong to a collective bargaining agreement corresponding to white zone, employers or relatives, active members of cooperatives and apprentices without link to the employer. Finally, it was only kept observations of workers with tenure (number of years working at the same firm) not surpassing 50 years and it was eliminated observations of workers working at firms in the agriculture, animal production, hunting, forestry or fishing sector (eliminate observations of workers in sector A according to *Classificação Portuguesa das Actividades Económicas Rev.3* (CAE Rev.3) or sector A and B according to CAE Rev.2.1) due to a lower presence of reported wage-earners and low coverage of the sector. Gross monthly earnings from dependent work are obtained by summing the earned remuneration of the worker and some irregular instalments that can be received too. In order to work with gross hourly labor earnings, it is divided the gross monthly earnings by the number of normal monthly hours of paid work. It was used the consumer price index (CPI)[15] deflators in the different years for Portugal to convert nominal hourly wages in real hourly wages. After treating the datasets on each year, the data is appended and it is formed a panel where each worker ID is tracked over time.

Education is accessed by the use of one-digit level for the highest education level completed. Although, there were changes on education categories at different years, it was still possible to aggregate individuals in each year in groups with: i) less than high school education, ii) high school education level and iii) college education or more, based on the variable of highest education level completed. Worker's experience is obtained through their age variable subtracting the years of education (it was subtracted 16 if they completed less than high school education level, 18 if they completed the high school education level and 21 if they completed the college education or more). However, the experience of a worker on the current firm can also be evaluated through tenure corresponding to the number of years working at the same firm.

The classification system used for worker's occupation in QP is *Classificação Nacional de Profissões* (CNP) in the initial years and *Classificação Portuguesa de Profissões* (CPP) in the last years. In particular, it is used CNP/94 before 2010 and CPP/2010 in 2010 and after. Thus,

[15] The CPI values were not part of QP and were obtained from OECD.

when occupations are required for the analysis, it is only used the data from 2010 onwards, as using correspondence tables between CNP/94 and CPP/2010 to harmonize the classification is not technically feasible, unless it is used the four-digit level of occupations. The one-digit level classification used in 2010 onwards is given by the major groups from CPP/2010, having: 0) armed forces occupations, 1) representatives of legislative and executive bodies, officers, directors and executive managers, 2) specialists in intellectual and scientific activities, 3) technicians and intermediate occupations, 4) administrative staff, 5) personal service, protection and security workers and salespeople, 6) farmers and skilled workers in agriculture, fisheries and forestry, 7) skilled workers in industry, construction and craftsmen, 8) plant and machine operators and assembly workers, 9) unskilled workers and 10) workers with no occupation assigned.

The classification system used for firm's economic sector in QP is *Classificação Portuguesa das Actividades Económicas* (CAE). In particular, before 2007 it is used CAE Rev.2.1 while in 2007 and after it is used CAE Rev.3. When is necessary to use sector data, it is only used data from 2007 onwards, since harmonizing the classification on the different years through correspondence tables is not technically feasible, unless it is used the four-digit level of classification (which is not available in some years). The one-digit level classification used for economic sectors since 2007 is given by the sections on CAE Rev.3, having: A) agriculture, animal production, hunting, forestry and fishing (sector eliminated before), B) extractive industries, C) manufacturing industries, D) electricity, gas, steam, hot and cold water and cold air, E) water collection, treatment and distribution; sanitation , waste management and depollution, F) construction, G) wholesale and retail trade; repair of motor vehicles and motorcycles, H) transport and storage, I) accommodation, catering and similar, J) information and communication activities, K) financial and insurance activities, L) real estate activities, M) consulting, scientific, technical and similar activities, N) administrative and support service activities, O) public administration and defence; compulsory social security, P) Education, Q) human health and social support activities, R) artistic, entertainment, sports and recreational activities, S) other service activities and U) activities of international organizations and other extra-territorial institutions (section T does not appear because QP exclude employers of domestic service workers and people producing for own consumption).

After treating these datasets, the 1st and 99th percentile of real hourly wages in each year were trimmed to eliminate outliers on the wage distribution.

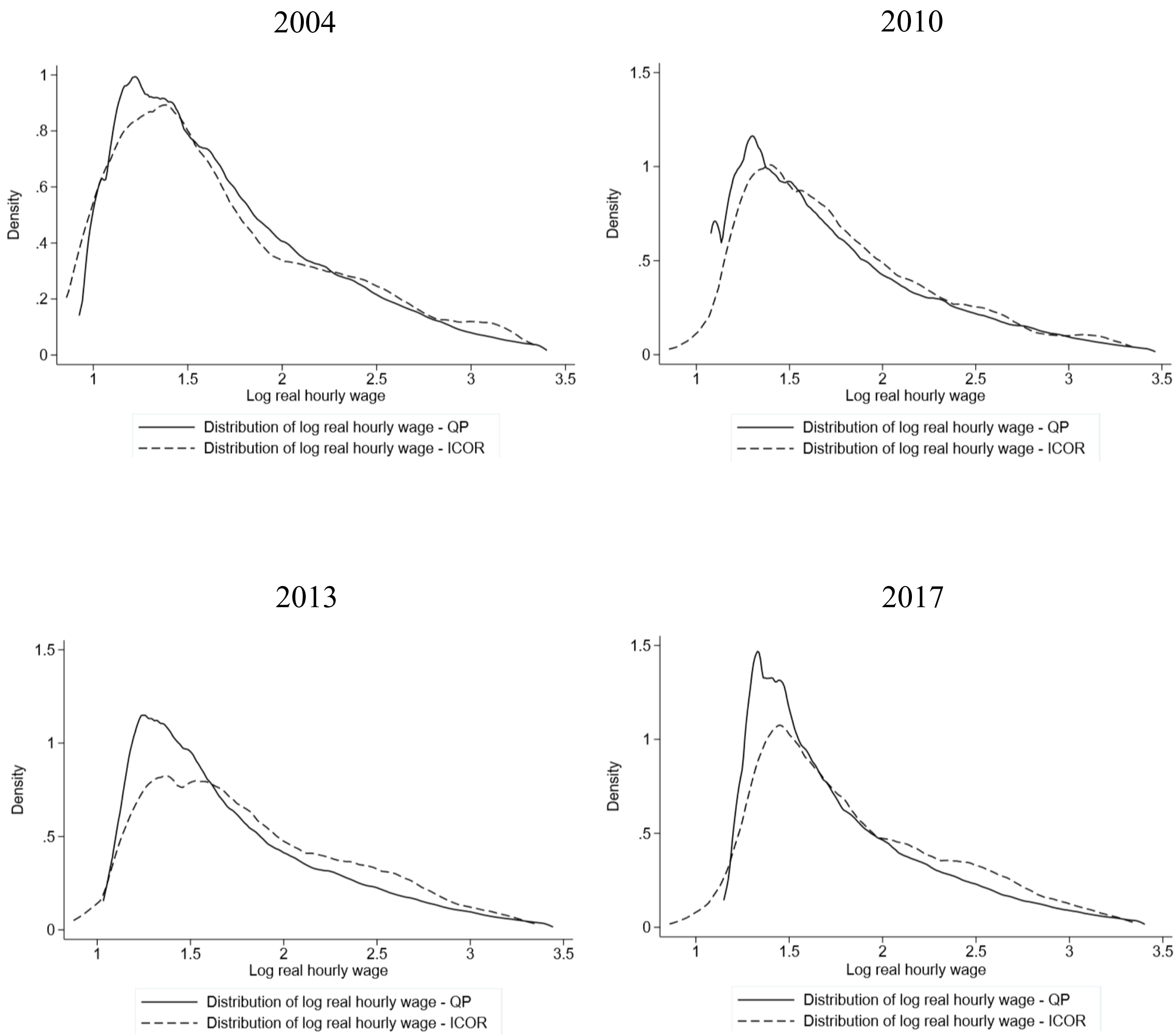
Table A1.1 compares, in each year, the number of full-time dependent workers and average log of real hourly wages on QP and ICOR and presents the number of firms on QP. Figure A1.1, following Kumler, Verhoogen, and Frias (2013), presents the cross-sectional estimates of wage distribution on QP and ICOR at 2004, 2010, 2013 and 2017, through Kernel density analysis of log of real hourly wages. The results indicate similarities between the wage distributions, even though log of real hourly wages is, on average, slightly higher on the survey and with Kernel density also slightly to the right. Even if the survey, by its nature, has much less worker's observations than the administrative census, these two datasets remain relatively comparable through time, having matching peaks in density.

Table A1.1 – Comparison of QP and ICOR in terms of average wage and size

Year	Number of fulltime dependent workers		Number of firms	Average log real hourly wage	
	QP	ICOR	QP	QP	ICOR
2004	1 630 826	3 541	208 186	1,681	1,686
2005	1 762 149	3 283	235 737	1,669	1,699
2006	1 714 284	3 120	224 857	1,676	1,685
2007	1 780 589	3 035	233 037	1,678	1,699
2008	1 814 234	2 743	238 522	1,690	1,767
2009	1 725 159	2 969	230 026	1,734	1,784
2010	1 771 585	3 165	218 059	1,732	1,785
2011	1 732 952	3 532	211 874	1,706	1,764
2012	1 583 340	4 074	192 277	1,686	1,773
2013	1 514 510	4 172	172 556	1,729	1,832
2014	1 531 052	4 382	172 284	1,746	1,810
2015	1 574 728	5 902	175 558	1,744	1,833
2016	1 602 240	7 390	176 916	1,754	1,844
2017	1 650 242	8 755	178 377	1,769	1,850
2018		10 391			1,966

Source: Computed using data from *Quadros de Pessoal* (QP) and *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Table informs about the number of full-time dependent workers between 18 and 65 years old and average log of real hourly wage in QP and ICOR and about the number of firms in QP in each year. Results in ICOR are generated using cross-sectional sample weights. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year and database were trimmed.

Figure A1.1 – Comparison of QP and ICOR in terms of wage distribution



Source: Computed using data from *Quadros de Pessoal* (QP) and *Inquérito às Condições de Vida e Rendimento* (ICOR)

Note: Figure provides a Kernel density analysis of the wage distributions from 2004, 2010, 2013 and 2017 in QP and ICOR. Wages appear as log of real hourly wages of full-time dependent workers between 18 and 65 years old in gross terms, using the consumer price index to convert in real terms. Results in ICOR are generated using cross-sectional sample weights. The 1st and 99th percentile of real hourly wages in each year and database were trimmed.

A2 – Additional tables and graphs

Table A2.1 – Percentiles of Portuguese wage distribution according to ICOR

Year	p5	p25	p50	p75	p95
2004	2,722	3,534	4,615	7,658	17,150
2005	2,734	3,645	4,739	7,499	17,147
2006	2,722	3,535	4,714	7,273	16,499
2007	2,781	3,549	4,831	7,294	16,721
2008	3,024	3,814	5,128	8,005	17,983
2009	3,117	3,855	5,231	8,263	17,721
2010	3,307	3,960	5,239	8,023	16,596
2011	3,223	3,844	5,055	8,068	16,221
2012	3,175	3,834	5,289	8,374	16,425
2013	3,227	4,076	5,612	9,019	17,336
2014	3,161	3,931	5,533	8,943	16,552
2015	3,322	4,116	5,576	9,025	16,833
2016	3,363	4,133	5,611	9,069	17,299
2017	3,443	4,261	5,522	8,815	17,200
2018	3,977	5,033	6,319	9,807	18,101

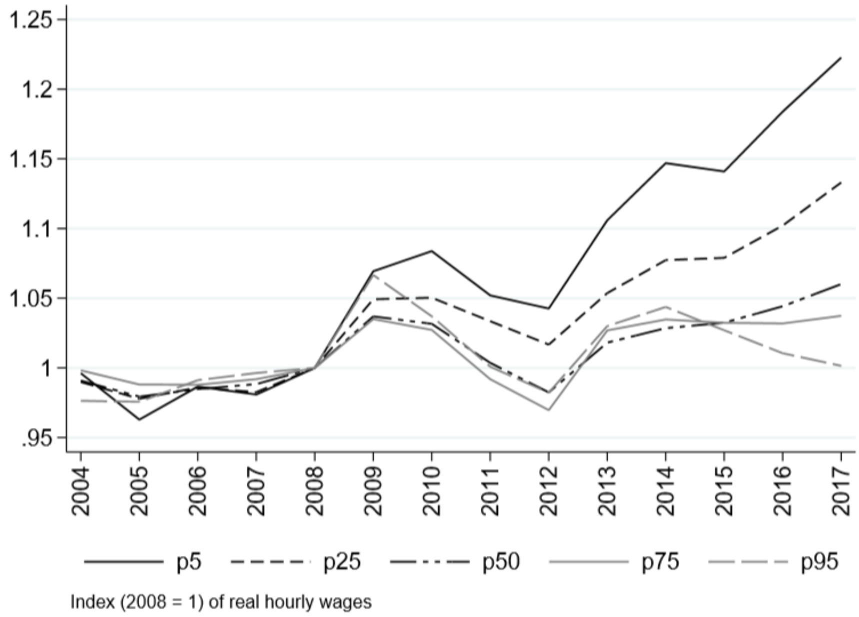
Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure provides the 5th, 25th, 50th, 75th and 95th percentiles of the real gross hourly wage distribution of full-time dependent workers between 18 and 65 years old for each year. Consumer price index is used to convert nominal wages in real terms. Results are generated using cross-sectional sample weights. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Table A2.2 – Percentiles of Portuguese wage distribution according to QP

Year	p5	p25	p50	p75	p95
2004	2,826	3,549	4,743	7,307	15,452
2005	2,732	3,506	4,688	7,232	15,441
2006	2,799	3,535	4,714	7,231	15,687
2007	2,783	3,523	4,731	7,260	15,768
2008	2,837	3,586	4,787	7,320	15,826
2009	3,034	3,762	4,964	7,576	16,881
2010	3,075	3,766	4,938	7,519	16,412
2011	2,985	3,706	4,804	7,260	15,839
2012	2,958	3,646	4,703	7,098	15,544
2013	3,138	3,778	4,874	7,516	16,301
2014	3,254	3,863	4,923	7,574	16,518
2015	3,237	3,869	4,942	7,556	16,254
2016	3,359	3,952	4,999	7,553	15,992
2017	3,469	4,063	5,075	7,593	15,849

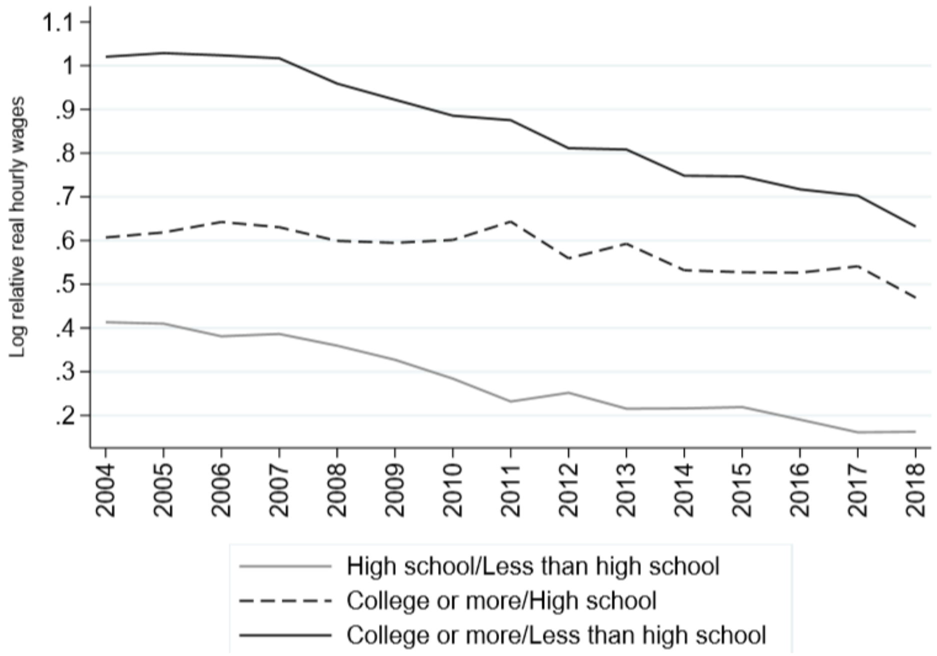
Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure provides the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of the real gross hourly wage distribution of full-time dependent workers between 18 and 65 years old for each year. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Figure A2.1 – Wage growth by percentile on QP



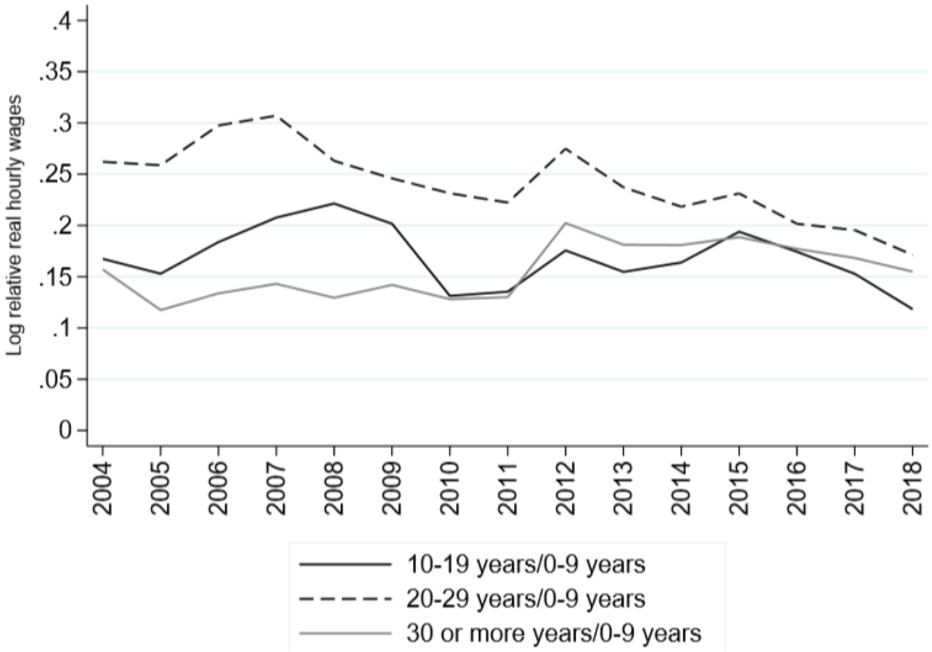
Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Figure presents the growth between 2004 and 2017 (index base at 2008) in the 5th, 25th, 50th, 75th and 95th wage percentiles. Wages appear as gross real hourly wages in main occupation and it was used the consumer price index to convert in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Figure A2.2 – Evolution of composition-adjusted education premium on ICOR



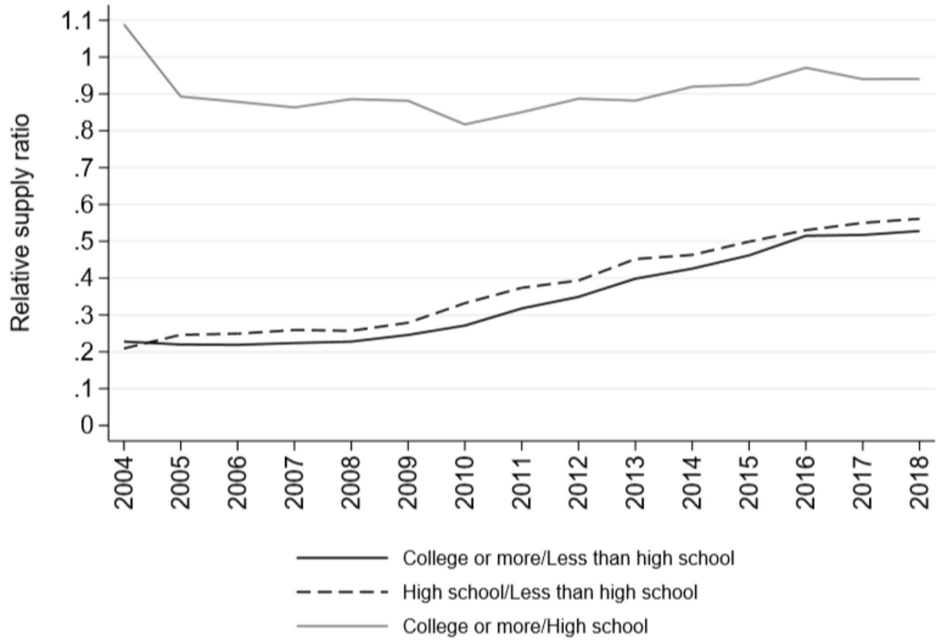
Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure presents composition-adjusted education premium between 2004 and 2018, in terms of log relative real hourly wages. Graph includes wage gap analysis for high school relative to less than high school level, college or more relative to high school level and college or more relative to less than high school level. The methodology used from building skill groups, average predicted log of real hourly wages by education groups and wage gaps to obtain the premiums on ICOR, is identical to the one used on QP. Education is divided in three categories based on worker's highest education level completed or still enrolled: college education or more (workers in which the highest education level completed was higher education), high school education (workers whose highest education level completed was high school or workers still enrolled at college level) and less than high school education (workers with primary education as the highest education level completed, workers still enrolled in high school education and workers in which primary education is incomplete or still enrolled). Worker's observations without information on education were eliminated. Experience of workers is given by the number of years of remunerated work and it is divided in four categories: 0-9 years, 10-19 years, 20-29 years and 30 or more years of experience. Worker's observations with no information on experience were eliminated. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. Results are generated using cross-sectional sample weights. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Figure A2.3 – Evolution of composition-adjusted experience premium on ICOR



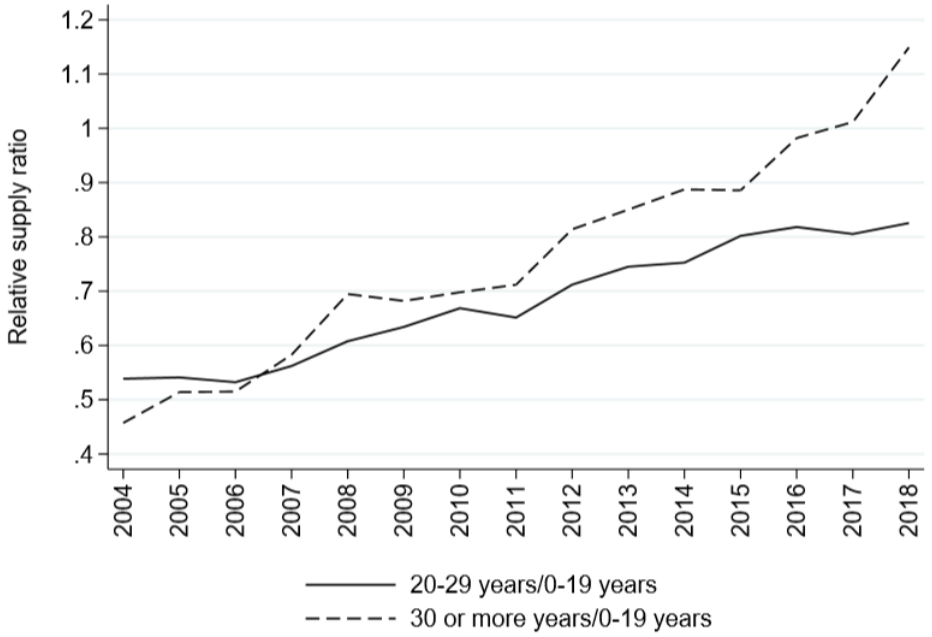
Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure presents composition-adjusted experience premium between 2004 and 2018, in terms of log relative real hourly wages. Graph includes wage gap analysis for workers with 10-19 years relative to 0-9 years of experience, 20-29 years relative to 0-9 years of experience and 30 or more years relative to 0-9 years of experience. The methodology used from building skill groups, average predicted log of real hourly wages by experience groups and wage gaps to obtain the premiums on ICOR, is identical to the one used on QP. Education is divided in three categories based on worker's highest education level completed or still enrolled: college education or more (workers in which the highest education level completed was higher education), high school education (workers whose highest education level completed was high school or workers still enrolled at college level) and less than high school education (workers with primary education as the highest education level completed, workers still enrolled in high school education and workers in which primary education is incomplete or still enrolled). Worker's observations without information on education were eliminated. Experience of workers is given by the number of years of remunerated work and it is divided in four categories: 0-9 years, 10-19 years, 20-29 years and 30 or more years of experience. Worker's observations with no information on experience were eliminated. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms. Results are generated using cross-sectional sample weights. The 1st and 99th percentile of real hourly wages in each year were trimmed.

Figure A2.4 – Evolution of relative labor supply of different education levels on ICOR



Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure presents relative supply between education groups from 2004 to 2018. Graph includes relative supply analysis of workers with high school relative to less than high school level, college or more relative to high school level and college or more relative to less than high school level. Education is divided in three categories based on worker's highest education level completed or still enrolled: college education or more (workers in which the highest education level completed was higher education), high school education (workers whose highest education level completed was high school or workers still enrolled at college level) and less than high school education (workers with primary education as the highest education level completed, workers still enrolled in high school education and workers in which primary education is incomplete or still enrolled). Worker's observations without information on education were eliminated. Worker's observations with no information on experience were eliminated. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms.

Figure A2.5 – Evolution of relative labor supply of different experience levels on ICOR



Source: Computed using data from *Inquérito às Condições de Vida e Rendimento* (ICOR)
Note: Figure presents the relative supply between experience groups from 2004 to 2018. Graph includes relative supply analysis of workers with 10-19 years relative to 0-9 years of experience, 20-29 years relative to 0-9 years of experience and 30 or more years relative to 0-9 years of experience. Worker's observations without information on education were eliminated. Experience of workers is given by the number of years of remunerated work and it was divided in four categories: 0-9 years, 10-19 years, 20-29 years and 30 or more years of experience. Worker's observations with no information on experience were eliminated. Wages appear as gross wages in main occupation and it was used the consumer price index to convert in real terms.

Table A2.3 - Decomposition of variance of wages across workers and firms

	2004 - 2009 (1)	2009 - 2013 (2)	2013 - 2017 (3)	Change 2004 - 2013 (4)	Change 2009 - 2017 (5)	Change 2004 - 2017 (6)
Variance of log of real hourly wages	0,2824 (100)	0,2671 (100)	0,2478 (100)	-0,0154 (100)	-0,0192 (100)	-0,0346 (100)
Variance of worker effects	0,1689 (59,8)	0,1861 (69,7)	0,1695 (68,4)	0,0172 (-111,8)	-0,0166 (86,1)	0,0006 (-1,8)
Variance of firm effects	0,0792 (28)	0,0637 (23,8)	0,0547 (22,1)	-0,0155 (101,1)	-0,0090 (46,8)	-0,0245 (70,9)
Variance of year effects	0,0011 (0,4)	0,0002 (0,1)	0,0010 (0,4)	-0,0009 (5,8)	0,0008 (-4,2)	-0,0001 (0,3)
2*Cov. worker and firm effects	0,0159 (5,6)	0,0025 (0,9)	0,0119 (4,8)	-0,0133 (86,9)	0,0094 (-49)	-0,0039 (11,3)
2*Cov. worker and year effects	-0,0007 (-0,2)	0,0001 (0,1)	-0,0008 (-0,3)	0,0008 (-5,5)	-0,0010 (4,9)	-0,0001 (0,3)
2*Cov. firm and year effects	0,0000 (0)	0,0001 (0)	-0,0001 (0)	0,0001 (-0,7)	-0,0001 (0,7)	0,0000 (0,1)
Residual	0,0181 (6,4)	0,0144 (5,4)	0,0116 (4,7)	-0,0037 (24,2)	-0,0028 (14,6)	-0,0065 (18,8)

Source: Computed using data from *Quadros de Pessoal* (QP)
Note: Table presents the decomposition of total variance of log real hourly wages, in each subperiod, in the variance of worker, firm and year effects, twice the covariance of worker and firm effects, twice the covariance of worker and year effects, twice the covariance of firm and year effects and variance of the residual (referred simply in the table as residual). Columns 1, 2 and 3 have the values and percentual contribution of each component to the variance of log real hourly wages, in each subperiod. Columns 4, 5 and 6 have the values and percentual contribution of each component to the change in overall variance. The percentual contribution of each component on the level and change of variance of log real hourly wages is presented inside parenthesis. The levels and change of each component are presented without parenthesis. The focus continues to be on gross wage's inequality/variance among full-time dependent workers in the main occupation. Consumer price index is used to convert nominal wages in real terms. The 1st and 99th percentile of real hourly wages in each year were trimmed.