The Role of Firms in Wage Inequality Dynamics*

Martim Leitão
Universidade Católica Portuguesa

Jaime Montana
Universidade Católica Portuguesa

Joana Silva
Universidade Católica Portuguesa, World Bank, and CEPR

July 1, 2022

Abstract

This paper examines the mechanisms through which firms impact earnings inequality dynamics. Using a rich combination of administrative matched employer-employee-job title data, and detailed technology adoption firm surveys for Portugal, we show that the decrease in wage inequality has arisen from a compression in the firm pay premium, the job title pay premium and their covariance. These effects were mainly driven by a decline in passthrough from firm characteristics to pay, rather than changes in the distribution of these characteristics. Results show that workforce composition and labor productivity are the main drivers of firm pay premiums compression, and that this effect comes from a decline in returns to these characteristics. An increasing share of workers earning the minimum wage and a reduction in labor market concentration also contributed to the fall in between-firm pay premium dispersion but had smaller roles. We also find that technological adoption increases within-firm labor income inequality. Our results shed new light on how firms impact labor income inequality dynamics and have profound policy implications for the design of policies to mitigate inequality.

Keywords: Labor Markets, Wage Inequality, Firms, Institutions, Technology, Concentration

JEL Classification: J00, J31, J40

*mmleitao@ucp.pt, jdoncel@ucp.pt, joana.silva@ucp.pt. We are grateful to Anna Bernard, Diana Bonfim, Ana Rute Cardoso, Isabel Horta Correia, Francisco Espiga, Ana Margarida Fernandes, Madalena Gaspar, Jaime Leyva, David Margolis, Hector Moreno, Pedro Raposo, Hugo Reis, Catarina Reis and participants in several seminars for helpful comments. We thank the Portuguese Institute of Statistics (INE) for data access. We remain responsible for any errors. We thank FCT for financial support. We declare that we have no relevant or material financial interest that relate to the research described in this paper.

1 Introduction

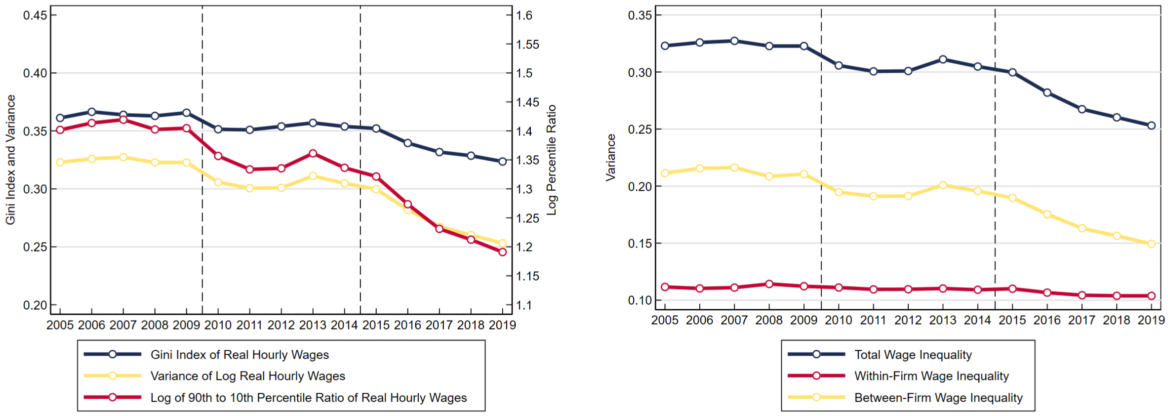
Rising earnings inequality is a widespread concern in developed and developing countries. Significant resources and a tremendous amount of effort have been devoted to fighting rising inequality. Commonly enacted policies to mitigate inequality generally target workers, ranging from education to policies to foster mobility and access to services. Fewer policy efforts have focused on firms and even fewer on the type of jobs workers perform. Hindering these efforts is a lack of understanding of how employers and job titles might mediate earnings inequality dynamics. Despite an extensive literature highlighting firms' contribution to changes in inequality (Card, Heining and Kline, 2013; Alvarez, Benguria, Engbom and Moser, 2018; Song, Price, Guvenen, Bloom and Von Wachter, 2019; Bonhomme, Lamadon and Manresa, 2019; Messina and Silva, 2021), the specific drivers and the channels through which they operate have not been measured. Quantifying them has the potential to help better calibrate inequality-mitigating policies, ultimately contributing to more efficient allocation of resources.

This paper fills this gap by exploring the mechanisms through which firms and job titles impact earnings inequality dynamics. We provide new evidence on how firm characteristics affect earnings inequality dynamics − through firms and job titles − by exploiting a large and persistent reduction in Portuguese wage inequality. Figure 1, panel (a) below shows the decline of Portuguese wage inequality over the twenty-first century (a decline of around 20 percent). Figure 1, panel (b) shows this same series decomposed into its within- and between-firm components.[1] The decline in inequality in Portugal was almost entirely driven by the decline in between-firm inequality, suggesting that differences in average pay across firms matter for wage dispersion. Previous literature has pointed out that dispersion in average pay differences across firms (between-firm inequality) could be driven by changes in firms' intrinsic characteristics or by changes in the sorting of good workers into good firms. For the United States, Song et al. (2019) find that the worker-specific component of pay, together with sorting, are responsible for the steady increase in earnings inequality, while firm pay premiums played no part. For West Germany, Card et al. (2013), find that roughly all components contribute to a steady increase in earnings inequality, although sorting and the worker component are more prominent. In contrast, Alvarez et al. (2018) find that dispersion in firm pay premiums is largely predominant in explaining the decline in earnings inequality in Brazil. Messina and Silva (2021) find a similar result for Latin American countries (Ecuador, Brazil, and Costa Rica).

In this paper, we show that the compression of firm pay premiums is almost single-handedly responsible for the sharp decline in earnings inequality observed in Portugal. In the absence of firm-specific effects, sorting would have driven up inequality, echoing the findings for Germany or the United States. The compression of job title pay premiums and changes in the sorting of good job titles into good firms were also important for this decline. The present paper investigates the channels underlying the compression of these three components. For all the components, we show that the decline in passthrough overcompensates changes in the distribution of firm characteristics and is entirely

[1] Overall inequality dynamics may stem from systematic differences in pay across firms (*between-firm inequality*) or from differences in pay within each firm (*within-firm inequality*) Card et al. (2013); Alvarez et al. (2018); Song et al. (2019); Messina and Silva (2021)

Figure 1: Portuguese Wage Inequality Dynamics (2005-19).



(a) (b)

Sources: *Quadros de Pessoal*, $2005-19$.
Note: Panel (a) plots three inequality measures, the Gini coefficient of real hourly wage, the variance of log real hourly wages, and the log ratio of real hourly wage at the 90^{th} and 10^{th} income percentiles, between 2005 and 2019. Real hourly wages are gross nominal wages of full time dependent workers deflated by the CPI (base=2015). The 1^{st} and 99^{th} percentiles of real hourly wages were trimmed every year. Panel (b) depicts the yearly evolution of the variance of hourly wages ("total wage inequality") for the 2005-2019 period, decomposed in a within firm inequality and a between firm inequality components. The vertical sum of within and between firm inequality components add up to overall inequality, for each year.

responsible for the decrease in earnings inequality. Taken together, labor productivity and firm knowledge composition are the main drivers behind the compression of firm pay premiums, and this effect comes from a decline in the returns to these characteristics (passthrough effect) rather than from compression of their distributions (composition effect).

While there is extensive literature recognizing the contribution of firm pay premiums to inequality (Alvarez et al., 2018; Song et al., 2019), the literature investigating the contribution of job titles to wage dispersion dynamics remains scant. In fact, a variety of papers model earnings as linearly decomposable into a worker-specific component and a firm-specific component, following Abowd et al. (1999) (henceforth, AKM), while omitting job titles. Yet, disregarding job titles while studying earnings inequality dynamics may not only compromise the robustness of our inferences, but also jeopardize our ability to determine the exact drivers behind changes in the wage distribution. Job titles matter for robust inference for several reasons. First, if they are important for wage-setting, their omission might jeopardize estimation of the firm and worker fixed-effect components through an omitted variable bias. Second, job titles matter since firm pay premiums and worker-specific characteristics may be correlated with job title fixed effects, and these covariances might thus matter for inequality. A third danger of omitting job titles is that if the error term is heteroskedastic (Andrews et al., 2008), the estimated variance explained by the pay premium and assortativity might be biased Kline et al. (2020) (henceforth, KSS) also thereby affecting the ability to carry out inference. As shown in previous studies, the KSS correction considerably reduces the sample size, potentially reducing wage variance and the contribution of the firm component to overall variance (Bonhomme et al., 2019; Kline et al., 2020).

We circumvent these obstacles by focusing on the unique context of Portugal, where rich administrative linked employer-employee job title data are available along with firms' financial account statements. Unlike a wide array of studies that have leveraged the AKM framework to isolate the role of firms in inequality, our data allow for the inclusion of job title fixed effects in our wage equation. In this setting, we extend the KSS leave-one-out methodology to the three high-dimensional fixed effects setting. By considering job titles, the loss in sample size when removing the articulation vertex (leave one out) sample is mitigated. This contributes to the robust estimation of firm pay premium dispersion, worker fixed effect dispersion, and job title fixed effect dispersion.

Guided by a simple search and matching model and by a set of stylized facts, we then provide the first comprehensive analysis of (i) the direction and magnitude of firm size, firm productivity, firm workforce composition, firm exposure to the minimum wage, and market concentration on firm pay premium dispersion and job title fixed effect dispersion; and (ii) the magnitude of the passthrough and composition effects associated with each of these characteristics. Specifically, we are interested in determining whether the observed decline in earnings inequality is driven by a weakening passthrough from firm characteristics to pay, or by firms becoming more homogeneous in their characteristics over time.

Our empirical analysis proceeds in two stages. In the first stage, we investigate the drivers of between-firm inequality through a two-step procedure. In a first step, we disentangle observed wage dynamics into the contributions of worker, firm, and job title heterogeneity, as well as their co-movement. Following Card et al. (2013), we provide evidence that the strong separability and exogenous mobility AKM assumptions are met in the data. In our first step, we initially do not control for worker and firm observable characteristics. Rather, closely following Alvarez et al. (2018)'s two-step procedure, in a second step, we explain firm fixed effects based on observable characteristics. We verify empirically that this two-step procedure is free of omitted variable bias. That is, in the second step, we inquire how observed firm characteristics translate into (i) the expected value of firm pay premiums, and (ii) the dispersion of these premiums. To evaluate how firm characteristics translate into firm fixed-effect dispersion, we project the non parametric variance counterpart, namely, the recentered influence function (RIF) (Firpo et al., 2009, 2018) into these covariates.

From our first step, we find that heterogeneity across workers (related to their fixed characteristics) is the strongest determinant of wage variance in levels, explaining between 44.01 and 51.55 percent of overall wage inequality. Yet, firms also play a key role in the level of wage inequality, with firm heterogeneity explaining between 18.78 and 21.82 percent of total wage variance. The association between high-earning workers and high-paying jobs accounts for between 8.38 and 10.85 percent of total wage dispersion. These results are consistent with the literature (see, for example, Portugal et al. (2018) for evidence for Portugal, and Alvarez et al. (2018) and Card et al. (2013) for evidence for Brazil and Germany, respectively). Overall, we find that the reduction in wage dispersion was due to a reduction in firm fixed-effect dispersion (around 60 percent), a decline in job title fixed effect dispersion (around 10 percent), and a decline in *firm-job assortativity* dispersion (around 7 percent).[2] Meanwhile, stagnation of the dispersion of worker fixed effects prevented a more substantial decline in overall inequality.

From our second step analysis of covariate projection onto firm fixed effects, we find that firm observables (such as size, concentration, and the share of workers earning the minimum wage) explain around one-third of the variability of firm fixed effects. We find that labor market concentration and the share of workers earning the minimum wage contribute negatively to the expected value of firm fixed effects, while workforce composition, labor productivity, and market concentration contribute positively, which is consistent with our conceptual framework. We also find that workforce composition contributes positively to the dispersion of the firm fixed effects, along with the share of minimum wage workers and product market concentration. Our Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973; Kitagawa, 1955) decomposition reveals that the decline in dispersion explained by these characteristics was driven by the reduction in the premiums of the characteristics rather than due to changes in the distribution of these characteristics. We show that had the returns to these characteristics not declined, the dispersion of firm fixed effects would have increased by 8.2 percent, which would have contributed to a 1.5 percent increase in wage inequality. We conclude that lower passthrough from firm characteristics to pay played a key role in compressing firm pay premiums. A key finding of our endeavor is that workforce composition and labor productivity are the main drivers of the compression of the firm pay premium dispersion. The decrease in the passthrough of value added per worker and firm knowledge composition, taken together, explain the largest share of the decrease in the variance of the firm fixed effects.

In the second stage, the paper explores the determinants of within-firm inequality. A wide array of factors, such as trade integration, demographic changes, or digital transformation, could simultaneously affect within-firm inequality. In this paper, we restrict our attention to the latter of these channels by leveraging a unique survey on the adoption – and diffusion – of information technologies that was carried out on a subset of Portuguese firms. The survey is unique not only because of its representativeness, but also because it can be merged with our administrative labor and financial statements data. Furthermore, the survey contains questions on firms' adoption of big data, allowing us to use a strong proxy for the elusive concept of technology. We rely on propensity score matching to estimate the causal effect of the adoption of big data on within-firm wage dispersion. Our within-firm analysis shows that firms' adoption of technology, in the form of big data, increases within-firm inequality by as much as 37 percent of the median of within-firm inequality. This result sides with empirical evidence showing that automation may exact a toll on inequality through skill-biased adoption or wage declines for workers specialized in routine tasks in sectors witnessing steep automation (Acemoglu and Restrepo, 2019).

Taken together, our findings have important implications for policymakers interested in mitigating inequality with minimal waste of resources. As educational attainment levels continue to rise steadily in developed economies, the marginal impact of educational policy investment decreased, driven by diminishing returns. Likewise, with minimum

[2]We understand *firm-job assortativity* as the tendency in the economy for better types of jobs to be present in firms which employers offer higher wages.

wage levels increasingly compressing the wage distribution, increasing them any further may not be feasible. Instead, our findings suggest that policies that limit product market concentration or propel idiosyncratic technology adoption for technologically laggard firms may be effective in tackling inequality.

Our research is related to four strands of the literature. The first is the literature investigating the role of firms in wage inequality dynamics. Studies such as Dunne et al. (2004), Sorkin (2017) and Mueller et al. (2017) document that firm characteristics matter for wage dispersion. Following Card et al. (2013), a number of studies have disentangled the sorting and firm pay premium components of between-firm wage inequality (Barth et al., 2016; Alvarez et al., 2018; Song et al., 2019; Messina and Silva, 2021). Among these studies, Alvarez et al. (2018) establish a link between value added and the level of firm pay premiums. They find that more productive firms pay significantly more even after controlling for sorting. Bloom et al. (2018) document a negative relationship between firm size and pay premium. Yet, these papers remain silent on the contribution of these characteristics to the *dispersion* of firm pay premiums. We contribute to this literature by examining the institutional and firm-level channels driving changes in firm and job title pay premium dispersion. To the best of our knowledge, we are the first to use re-centered influenced functions (RIF) to this end.

The second strand is the recent literature highlighting the importance of job title heterogeneity for wage formation. Job titles capture the institutional and task compensation heterogeneity of the roles and occupations inside the firm (Cardoso et al., 2016a; Portugal et al., 2018; Raposo et al., 2021). For example, Dustmann et al. (2009) provide evidence that different occupations are not distributed uniformly in the support of the wage distribution. Consequently, occupation types (that is, the skills used in the occupation) can drive different wage inequality dynamics. A similar finding for routine and non-routine tasks (cognitive or manual) explains the impact of job polarization on the distribution of wages (Goos and Manning, 2007). Another strand of the recent literature highlights that including a richer description of the types of jobs (tasks and skills) could shed light on the dynamics of wage inequality (Autor and Handel, 2013; Acemoglu and Restrepo, 2019). Yet, to the best of our knowledge, none of these papers explores the relationship between firm characteristics and job title pay premium dispersion and their impact on earnings inequality.

Third, the first part of our empirical analysis relates to an extensive literature using large administrative data to decompose wages into their worker and firm heterogeneity components (Abowd and Kramarz, 1999; Cardoso et al., 2016b; Bloom et al., 2018; Alvarez et al., 2018). While most of the literature includes only worker and firm fixed effects in AKM wage regressions, a smaller but growing literature includes job title fixed effects as well (Carneiro et al., 2012; Raposo et al., 2021). In our baseline AKM regression, we thus consider the role of job title heterogeneity in wage determination.

Lastly, recent literature has pointed out the challenges in precisely decomposing the variance in wages. Our study relates to the literature on robust identification and inference of models with high dimensional fixed effects Andrews et al. (2008); Kline et al. (2020). A robust estimation could entail a reduction in wage dispersion, and the importance of the firm pay premium component (Bonhomme et al., 2019; Kline et al., 2020). In this paper, we extend the KSS correction to include worker, firm, and job title fixed effects. We show that incorporating a third high-dimensional fixed effect improves the connectivity of the network, and assures robust estimation.

The remainder of the paper is organized as follows. Section 2 lays down a conceptual framework as a motivation for our empirical analysis. Section 3 describes our main data sources and presents descriptive statistics. Section 4 presents a set of stylized facts that suggest that firms and job titles are major candidates behind the fall in wage inequality. Section 5 explores the drivers of between firm inequality dynamics. Section 6 explores the drivers of within firm inequality dynamics. Section 7 concludes.

2 Conceptual Framework

This section outlines the conceptual framework that guides our empirical approach. We rely on the canonical AKM two-way fixed effects model (Abowd et al., 1999), in which the log wage is additive on worker, firm, and job title fixed effects. Like Alvarez et al. (2018), we estimate a *'restricted'* wage model, of the following form:

$$w_{it} = \alpha_i + \psi_{j(i,t)} + \phi_{k(i,t)} + \tau_t + \epsilon_{it} \tag{1}$$

where α_i is a worker effect, which captures the time-invariant unobserved characteristics of each worker; $\psi_{j(i,t)}$ is a firm effect, which captures the firm pay premium component; $\phi_{k(i,t)}$ is a job title fixed effect which captures the time invariant unobserved characteristics of the different occupations; τ_t is a time fixed effect; and ϵ_{it} is an error term component. The error component is assumed to follow a conditional mean-zero assumption (see equation 2), ruling out worker mobility because of the error component. We follow Card, Heining and Kline (2013) to test empirically whether that is the case in the data, and that the gains of switching workers across quartiles of the distribution are symmetric.[3]

$$\mathbb{E}[\epsilon_{it}|\alpha_i, \psi_{j(i,t)}, \phi_{k(i,t)}, \tau_t] = 0 \tag{2}$$

The model is said to be *'restricted'*, since it does not incorporate observable time varying characteristics. Our aim is to understand the determinants of the non-varying components, following Alvarez et al. (2018), Song et al. (2019), and Messina and Silva (2021). In a later stage, we project the estimated fixed effects into time-varying characteristics. Our choice of which covariates to include in such projection is guided by the results of the wage-setting mechanism proposed in our theoretical model.[4]

$$\psi_{j(i,t)} = \beta_t \frac{z_j f(k_j)}{(1+\gamma_j)} \tag{3}$$

[3]The detailed empirical test of the exougenous mobility assumption can be found in section A.2, in the appendix.

[4]The model builds on insights from a simple Diamond-Mortensen-Pissarides model (Mortensen and Pissarides, 1994), with two-sided worker-firm heterogeneity (Postel–Vinay and Robin, 2002; Cahuc et al., 2006). We provide a detailed derivation of our model in an online supplemental appendix, while the main text presents its key insights. Such setting pins down the role of firms as key drivers of wage inequality dynamics.

Following our model results, the firm pay premium is directly related to several components: (i) total factor productivity, on which the firm-specific production technology is dependent. It is also a function of (ii) the worker bargaining weight β_t, which is assumed to be constant across workers and firms, but that we allow to change over time. The firm specific component also depends on (iii) a labor cost friction. The higher is the friction γ_j, the lower is the firm specific component $\psi_{j(i,t)}$. Finally, the firm-specific component depends on (iv) the capital stock. The capital stock is a proxy for the firm-specific skill requirements in production. A firm that invests a large amount of capital will require different types of workers for the different tasks and the contents of the jobs. After estimating the reduced AKM specification (equation 1), we decompose the wage variance into the worker, firm and job title components. In doing so, we explicitly evaluate the contribution of firms' heterogeneity to wage inequality (captured by the variance in wages), as well as the heterogeneity of workers and job titles. Concretely, for each subperiod, the wage variance is linearly decomposed as:

$$Var(w_{it}) = Var(\alpha_i) + Var(\psi_{j(i,t)}) + Var(\phi_{k(i,t)}) + Var(\tau_t) + Var(\epsilon_{it}) + \mathbf{2C}^T \quad (4)$$

where $Var(w_{it})$ stands for the variance of log real hourly wages. $Var(\alpha_i)$, $Var(\psi_{j(i,t)})$, $Var(\phi_{k(i,t)})$, and $Var(\tau_t)$ are the variances of worker, firm, job title, and year fixed effects, respectively. These variances represent, respectively, wage heterogeneity across workers related to their fixed characteristics, wage heterogeneity across firms from their distinct pay premiums and wage heterogeneity across different job-titles. The term $Var(\epsilon_{it})$ stands for the variance of the error term. The term $\mathbf{2C}^T$ is a 1 by 6 vector including the covariance between all combinations of the terms on the right-hand side of equation (1).

In levels, inequality in wages stems from inequality across firms in technology, capital intensiveness, and labor cost frictions. It also stems from heterogeneity in worker characteristics, causing divergence in their outside options. Finally, inequality also depends on the sorting of workers between firms and job-titles. Which of these effects makes a stronger contribution to overall wage inequality in period t depends on the relative bargaining power of the agents. If workers have full bargaining power, then $\beta_t = 1$ so that dispersion in firm heterogeneity plays a large role in explaining the level of wage inequality. How the variance of firm-specific characteristics translates into wage inequality depends both on a composition and on a passthrough component.

$$Var(\psi_{j(i,t)}) = \underbrace{(\beta_t)^2}_{Passtrhough} \times \underbrace{Var\left(\frac{z_j f(k_j)}{1+\gamma_j}\right)}_{Composition}$$

The change in inequality over time can be assessed by evaluating the change in the variance of wages across consecutive periods.

$$\Delta Var(w_{it}) = \Delta Var(\alpha_i) + \Delta Var(\psi_{j(i,t)}) + \Delta Var(\phi_{k(i,t)}) + \Delta Var(\tau_t) + \Delta Var(\epsilon_{it}) + \mathbf{\Delta 2C}^T$$

In dynamics, as long as there are changes in bargaining power over time, there will be changes in inequality over time. Concretely, if the bargaining power of workers increases over two consecutive time periods, so that $\Delta\beta_t^2 > 0$, then a smaller dispersion of firm characteristics leads to less inequality. Moreover, for the same dispersion of firm characteristics, a decrease in workers' bargaining power will lead to a decrease in dispersion of the firm pay premium, via lower passthrough from the characteristics into the pay premium.

3 Data and Descriptive Statistics

This paper's empirical analysis draws on three main datasets.

Quadros de Pessoal (QP): First, we use *Quadros de Pessoal* (QP), a matched employer-employee dataset collected by the Portuguese Ministry of Employment. QP covers and follows over time virtually all Portuguese private sector workers and firms with more than one worker, having close to 300,000 firms and more than 2.5 million worker observations each year over 2005-19. We restrict the analysis to full-time dependent workers between ages 18 and 65 years (working-age population). The dataset provides comprehensive information on workers' demographic characteristics (age, gender, schooling, and so forth), and job characteristics (occupational group, professional category, wage, hours worked, firm tenure, and so forth). For each worker, the employing firm is uniquely identified through a firm identification code. The firm-level characteristics in QP include, among others, sales, number of employees, equity, percentage of foreign capital, geographical location and date of creation, and the industry code according to the Classificação Portuguesa das Atividades Económicas (CAE).[5] These data have been recently used by Carneiro et al. (2012), Card et al. (2016), Card and Cardoso (2021), Raposo et al. (2021), and Carneiro et al. (2022).

Table 1 displays descriptive statistics for selected variables and indicators, by subperiods and for the overall period. The sample contains data on individual workers for which a fixed effect was estimated. The table presents statistics both for the largest connected set Abowd et al. (2002) [6], and for the *leave-one-out sample* (Kline, Saggio and Sølvsten, 2020). Our definition of the leave-one-out sample extends the KSS methodology to include worker, firm, and job title fixed effects. The three-way leave-one-out sample is constructed under a simple assumption on the network structure: workers are connected to firms, and firms are connected to job titles. We rule out the less restrictive possibility that workers connect directly to both firms and job titles, since that would decrease the probability of finding articulation points in the network. A first interesting result from table 1 is that the *three-way leave-one-out sample* is similar in size to the largest connected set. This result is in contrast to previous evidence, in which a two-way leave-one-out sample results in a large decrease in sample size, changes the estimated variance components and affects inference (Bonhomme et al., 2019; Kline et al., 2020). Extending to three components allows us to preserve workers who were not moving across firms but have similar job titles, sharing the institutional framework and job conditions that keep them connected to the network.

[5]Given that the Portuguese classification of firm's economic sectors (Classificação Portuguesa das Atividades Económicas - CAE) has been revised in 2007 to match the NACE Revision 2, a concordance was put together to ensure harmonization with the sectors' classifications from previous years.

[6]The largest connected set gives the largest sample in which all firms and job titles are connected by worker mobility. In support of the mobility assumption we present the mobility across firms and job titles in table A1, in the appendix.

Table 1: Summary statistics

Sample	2005-2009	2010-2014	2015-2019	2005-2019
Largest Connected Set				
N. Observations	8161585	7668852	7881089	26516202
Number of Firms	294550	247627	232870	481992
Number of workers	2254434	2061354	2145383	3684974
Number of Job-titles	34596	52696	43247	82015
Number of movers across firms	398716	297777	392864	1495186
Number of movers across job-titles	931510	1405720	1015331	2798628
Mean $\log(w)$	1.7092	1.7395	1.7827	1.7328
Variance $\log(w)$	0.3280	0.3084	0.2764	0.3049
Leave one out sample (KSS)				
N. Observations	8146049	7655728	7864707	26502639
Number of Firms	293682	246907	232304	481274
Number of workers	2241581	2051065	2135020	3672511
Number of Job-titles	34344	52275	42850	81913
Number of movers across firms	398288	297410	392589	1494914
Number of movers across job-titles	931059	1405275	1014265	2798429
Mean $\log(w)$	1.7095	1.7396	1.7823	1.7329
Variance $\log(w)$	0.3280	0.3084	0.2764	0.3049

Sources: *Quadros de Pessoal*, $2005-19$.

Note: The table displays descriptive statistics on the number of movers and units. The top of the table presents statistics for the largest connected set (three-way), which gives the largest sample in which all firms and job titles are connected by worker mobility. The bottom of the table displays statistics for the *three-way leave out sample*, which is the largest connected set such that every firm and job-title remains connected after removing any single worker from the sample. The first three columns present the summary information for each subperiod, while the last column present the key descriptive statistics for the whole sample over 2005-19.

Sistema de Contas Integradas das Empresas (SCIE): We also use *Sistema de Contas Integradas das Empresas* (SCIE), which is longitudinal, firm-level data set collected by Statistics Portugal (INE). This dataset links with QP through the unique firm identification code. SCIE covers all firms (companies, individual entrepreneurs, and self-employed) that produce goods or services during the year, excluding firms in the insurance and financial sector, those that produce agricultural products or entities that are not market oriented. From 2005 to 2019, each year has more than 1 million firm observations detailing their economic activity (for example, CAE industry code, geographical location (according to the Nomenclatura das Unidades Territoriais para Fins Estatísticos, NUTS, II), birth/death, and number of workers) and accounting statements. Generically, the dataset includes information on financing and accounting variables. Employment and labor productivity variables can also be extracted from SCIE.

Inquérito à Utilização de Tecnologias da Informação e da Comunicação das Empresa (IUTICE): In Section 6 we make extensive use of the *Inquérito à Utilização de Tecnologias da Informação e da Comunicação das Empresa* (IUTICE), an annual survey of a sample of firms having more than 250 workers (or firms having sales greater than 25 million €) on the adoption and diffusion of information technologies in firms. The survey includes, among others, questions on the use of internet and computers, and, for 2016 and 2018, questions on the adoption of big data. Through the firm key identifier, we can merge QP and SCIE firm information with IUTICE data.

4 Evolution of Earning Inequality in Portugal

Wage inequality in Portugal declined continuously over the course of the twenty-first century, by a staggering 20 percent. This fall in wage inequality does not depend on the measure chosen, and both the Gini, the log ratio and variance fell during this period. This decline in inequality was characterized by:[7]

1. The decrease in inequality affected the overall support of the income distribution, but the decrease was larger in the lower end of the distribution (see figure A8).

2. When we decompose wage dispersion into its within-firm component and its between-firm component (see figure 1), the latter is the most important in levels (around 60 percent). Moreover, the between-firm component decline during the period 2015-19 accounts for the bulk of the wage inequality decline over that time span. Such pattern is present also across sectors (see figure A1).

3. When we decompose the variance of wages into a within-skill and between-skill component, the within-skill is the largest one in levels (see figure A9). Even if both components decreased, the change in the between skill component is larger, and between-skill inequality is thus driving the fall in overall wage variance. This highlights the relevance of having a variable capturing different skills across demographic groups, such as job titles. This finding stands in contrast to what is found in the United States, where the dispersion in wage had occurred within-skill, resulting in an increase in inequality (Autor et al., 2020).

[7] See appendix B for a detailed explanation on the unambiguous decline of labor income inequality in Portugal, and a detailed description of the characteristics presented below.

4. Historically, large firms have paid significantly higher wages. When we analyze the evolution of this relationship trough time, we find that the relation between firm size and wages in Portugal has weaken over time considerably (see figure A10).

Taken together, these stylized facts point towards the fundamental role of firm and skill heterogeneity for wage determination. However, different firms may have very different wage profiles due to systematic differences in the workers they hire or type of jobs carried (Gerard et al., 2021). Likewise, similar skills might be rewarded differently across firms and these initial within-between mechanical decompositions do not capture these effects jointly.

Table 2 addresses this concern by presenting the AKM variance decomposition for Portugal following equation 4 for the three subperiods considered and the whole sample. The worker effects are the most important source of heterogeneity, followed by the firm effects and sorting between workers and job titles. Even if the job title heterogeneity is not as large as the worker or firm heterogeneity, its size is comparable to that of worker-firm sorting. These results are also consistent for the sample as a whole. Interestingly, for the whole sample, the worker variance decreases, and worker-firm sorting increases.

Table 3 presents the changes in composition throughout the samples. The last two columns show the changes in the second and third sub-samples with respect to the first period. The last column in table 3 shows that the negative trend in inequality is totally explained by a decrease in the variance of the firm effect component, the job title, and the covariance of firm and job title effects. These findings add on to the literature highlighting the role of firms as major actors in wage inequality dynamics. For Brazil, Alvarez et al. (2018) find that the firm component explains around 39 percent of the decline in wage inequality between 1996 and 2012. In Germany and the United States, the change was led by better workers sorting into better firms (Card et al., 2013; Song et al., 2019). In the next section, we pin down and quantify the channels underlying the compression of the firm effect, of the job title fixed effect, and of the covariance of firm and job title effect.

5 Firm Pay Premium and Firm Characteristics

In this section, after having *robustly* estimated the restricted AKM (Abowd et al., 1999), we project the estimated fixed-effects onto covariates, and later extract the passthrough to the firm fixed-effects variance. Specifically, we project two outcomes. First, we project the fixed-effects. This corresponds to explaining how the covariates change the level (expected first moment) of the firm-specific pay premium distribution. Second, we project the non-parametric variance counterpart, namely, the RIF - recentered influence function - (Firpo et al., 2009, 2018). This aims to explain how the covariates change the dispersion (the second moment of the distribution). Finally, we decompose changes in the firm pay premium dispersion dynamics into a *return effect* and a *composition effect*. That is, we disentangle the part of the change in the firm pay premium dispersion that comes from changes in the composition of the characteristics of pay, and the part that comes from changes in the returns of those characteristics.

Table 2: Wage variance decomposition - AKM

Sample	2005-2009		2010-2014		2015-2019		2005-2019	
	Value	Share (%)	Value	Share (%)	Value	Share (%)	Value	Share (%)
Variance $\log(w)$								
Plug In	0.3280		0.3084		0.2764		0.3049	
Leave out (KSS)	0.3279		0.3083		0.2761		0.3049	
Variance workers effects								
Plug In	0.1443	44.01	0.1590	51.55	0.1330	48.11	0.0989	32.45
Leave out (KSS)	0.1444	44.02	0.1590	51.58	0.1330	48.17	0.0989	32.44
Variance Firms effects								
Plug In	0.0820	24.99	0.0673	21.82	0.0519	18.78	0.0589	19.32
Leave out (KSS)	0.0819	24.99	0.0673	21.82	0.0519	18.79	0.0589	19.32
Variance Job-title effects								
Plug In	0.0190	5.81	0.0163	5.30	0.0136	4.91	0.0212	6.97
Leave out (KSS)	0.0190	5.80	0.0163	5.29	0.0135	4.89	0.0213	6.97
Covariance of Worker.Firm ($2\times$)								
Plug In	0.0163	4.98	0.0087	2.83	0.0206	7.44	0.0342	11.22
Leave out (KSS)	0.0164	4.99	0.0087	2.83	0.0205	7.42	0.0342	11.22
Covariance of Worker.Job-title ($2\times$)								
Plug In	0.0336	10.24	0.0258	8.38	0.0300	10.85	0.0441	14.47
Leave out (KSS)	0.0336	10.23	0.0258	8.36	0.0299	10.83	0.0441	14.47
Covariance of Firm.Job-title ($2\times$)								
Plug In	0.0125	3.81	0.0149	4.84	0.0103	3.73	0.0177	5.81
Leave out (KSS)	0.0125	3.80	0.0149	4.83	0.0102	3.69	0.0177	5.81
Coeficient of determination R^2								
Plug In	0.9137		0.9254		0.9161			0.8892
Leave out (KSS)	0.9137		0.9254		0.9159			0.8893

Sources: *Quadros de Pessoal*, $2005-2019$.
Note: The table displays the labor income variance decomposition for the three different sub-periods in consideration and the whole sample (left two columns). The variance decomposition follows equation 4 in the text on the restricted AKM especification (equation 1). The decomposition is performed in the largest connected set and in the three-way leave one out sample, which is the extension of KSS method for the three high dimensional fixed effects case.

Table 3: Changes in the Composition of Wage Variance - Largest Connected Set

	2005-2009	2010-2014	2015-2019	$\Delta_{1.2}$	$\Delta_{1.3}$
	(%)	(%)	(%)		
Variance workers effects	44.01	51.55	48.11	7.542	4.099
Variance Firms effects	24.99	21.82	18.78	-3.170	-6.211
Variance Job-title effects	5.81	5.30	4.91	-0.508	-0.900
Covariance of Worker.Firm (2×)	4.98	2.83	7.44	-2.154	2.454
Covariance of Worker.Job-title (2×)	10.24	8.38	10.85	-1.861	0.610
Covariance of Firm.Job-title (2×)	3.81	4.84	3.73	1.030	-0.084

Sources: *Quadros de Pessoal*, $2005-19$.
Note: The table displays the share of variance explained by each component in the largest connected set sample. The last two columns present the changes between the the second and third sub-periods with respect to the first one.

Many channels could drive the firm-specific pay premium and its dispersion. Innumerable factors could affect firm performance and wage-setting behavior, and choosing them a priori would be incorrect. Our choice of the channels that affect the firm pay premium and firm performance depends on five elements: the determinants of the firm wage component laid down in Section 2, our stylized facts, the evidence presented in previous economic literature, the Portuguese institutional setting during this period, and finally data availability. To capture changes in the firm-specific pay premium, we thus take into consideration the effects of size, firm performance, market power, and institutional settings.

Firm Size Following our stylized fact reporting the decline in the large firm pay premium, we acknowledge that one of the main firm characteristics that explains the firm pay premium is firm size. The role of firm size has been analyzed in the case of Brazil (Alvarez et al., 2018). Previous literature highlights that working conditions improve with firm size both in monetary and non-monetary compensation (Bloom et al., 2018). To explain how firm size is attached to pay premiums, several explanations have been advanced. First, rents might have a role to play, and size might relate to rent splitting hence affecting wages. Given that larger firms tend to have larger rents, when those rents are split among workers, the share is larger for workers in larger firms. Such difference is larger even if the splitting rule is constant across firms, which might not be the case since for larger firms, the chances of workers coordinating and having greater negotiation power might accentuate this channel, further increasing wages. Another explanation is that larger firms pay efficiency wages to maintain and attract more productive workers. This being the case, the average wage is higher in larger firms. Finally, another explanation is that the environment in a larger firm is less agreeable, so firms must compensate workers for this (*compensating differentials*). In this paper, we use employment within the firm as our measure of firm size.

Firm Performance The role of firm performance naturally comes to mind when proposing determinants for the firm-specific pay premium. Changes in productivity directly affect the match surplus value for the firm, which is translated directly into the firm specific component and then to wages (see equation 3). Capturing differences in firm performance at the firm level is known to be challenging. To capture firm performance we use two proxies: *value added per worker*, and *workforce composition*. On the one hand, value added per worker is generally used as a proxy for productivity. Depending on the level of productivity the firm might or might not share rents with workers (Card et al., 2016)[8]. On the other hand, technology adoption changes the nature of the tasks and skills required by the firm. This has an impact on how the firm is organized, in terms of occupations. To capture the organizational dimension of the firm, we consider workforce composition (Acemoglu and Restrepo (2020)), as a proxy for the firm's technology adoption. We propose a skill index that encompasses the skill average requirements of the workforce, which is a proxy for the knowledge composition of the firm. To build this measure, we start by adding a series of O*NET skill requirements information on each occupation, as captured by our harmonized classification of occupations (according to the *Classificação Nacional das Profissões*). This vector of skill requirements encompasses thirty-five characteristics (for example, active learning levels, intensity of complex thinking required, mathematics needed and so forth). We then reduce this large set of descriptors to a single dimension using Principal Component Analysis (PCA) on this set of O*NET measures. Finally, we normalize this index so that it takes a value bounded between zero and one. We expect that firms that are more intensive in technology investments require and employ workers with higher skills, impacting the firm's organization and composition. Changes in productivity, or changes in the workforce composition of the firm, might naturally affect the firm compensation policy, and have effects on labor income inequality (Acemoglu and Restrepo, 2019, 2020, 2021).

Market Concentration The role of imperfect market competition in labor market outcomes has been of great interest lately, and there is evidence that labor market power is depressing wages (Naidu et al., 2018; Naidu and Posner, 2021; Azar et al., 2019). When the relative size of the firm is larger with respect to the market in which it operates, the firm might have greater power in wage-setting negotiations, resulting in lower worker wages. This might happen because the lack of firm competition undermines the credibility of an outside option for workers[9] or because firms could coordinate among themselves, thus impacting workers' earnings. To capture such behavior, we use two proxies for imperfect market competition. First, we use a proxy for market industry concentration using the average Herfindahl–Hirschman Index (HHI) computed using firm sales at the 4-digit economic sector level. Second, we use another measure that accounts for labor market concentration. To capture this, we calculate the (HHI) using the employment share for each local labor market, defined as the 4-digit occupation level at the 2-digit regional level.[10]

[8]We use Card et al. (2016) definition of value added per worker, where we assign 0 to the value added per worker level that do not share rents.

[9]Consider the case of two firms operating in the same market. Both firms hire the same type of workers, and if the firm is large enough compared to its competitor, the claim of an outside option from a worker is not credible. In this case the worker has less wage-negotiation power.

[10]This local labor market definition allow us to compare a specific occupation (i.e. secretaries) in a region (i.e. Algarve).

Institutional Factors Section 4 provides evidence on how different parts of the support of the distribution changed during the period of analysis. This suggests that different demographic/occupational groups have been affected diversely. Aside from the previously described dimensions, and ruling out sorting between the fixed effects, we consider institutional factors that could explain why the decrease in inequality was driven by the lower tail of the earnings distribution. Among the candidates is the institutional framework for wage bargaining in Portugal, which is composed of a national minimum wage and a set of collective bargaining agreements. The importance of the minimum wage has been widely discussed in economics, and Leung (2021) points out how changes in the wage might affect real wage inequality differently in the United States. This channel has also been considered for the case of Portugal (Portugal and Cardoso, 2006). In the same fashion, the role of collective agreements in wages and employment has been the object of discussion. There is evidence of its importance across countries (see Devicienti et al. (2019) and Fanfani (2020) for the case of Italy, and Card and Cardoso (2021) for the case of Portugal). To our knowledge, there is no literature directly related to the impact of those channels in inequality dynamics. In our desired specification, equation 1, the job title fixed-effects capture the heterogeneity in collective agreements. This papers considers the share of minimum wage workers in the firm and evaluates how this characteristic might impact pay premium dispersion and its dynamics.

The following subsections provide explanations for how size, firm performance, market concentration and the minimum wage decrease the dispersion of the identified components —firm, job title, and their covariance —, pushing down wage inequality. We want to understand how such factors, in levels and dispersion, affect wage inequality, which will then allow us to quantify their overall contributions.

5.1 First Moment of the Fixed Effect Distribution

Following Alvarez et al. (2018)'s approach we analyze firms' observable characteristics that might influence their wage rates. To do so, we project the firm fixed effects on the above covariates. We also do the exercise for job title fixed effects, and for the covariance of the firm and job title fixed effects. This will determine the factors that influence the fixed effects on average. For each sub-period P, using firm observations, and weighting by the number of worker-year observations, the following model is run by OLS:

$$\hat{y}_j^P = \alpha^P + \bar{\mathbf{X}}_j^P \Gamma^P + s_j^P + r_j^P + \mu_j^P \tag{5}$$

where $\hat{y}_f$ stands for the estimated firm or job-title fixed-effects in each of the sub-periods. $\bar{\mathbf{X}}_j$ is a matrix of the average firm characteristics for firm j in each subperiod, and μ_f stands for an i.i.d. error term. The matrix $\bar{\mathbf{X}}_j$ includes the average size, value-added per worker, industry concentration, labor market concentration, workforce composition and share of workers at the minimum wage. α is the subperiod-specific regression intercept. To proxy value added per worker, we use the average log of gross value-added (at market prices) during the sub-period, while to proxy industry concentration we use the average Herfindahl–Hirschman Index (HHI) computed at the 4-digit economic sector level, for each subperiod. Our specification also accounts for firm sector of activity (at the 2-digit level) fixed effects and firm region (at NUTS II) fixed effects.

In all the regressions, the estimated fixed effects have been re-scaled with respect to the largest firm during the first sub-period, such that this firm fixed effect is zero. We perform this re-scaling to compare firm fixed effects levels over time. We take advantage of the Julia implementation of the Correia method for estimating high-dimensional categorical variables.[11] The method assigns the value of zero to the base category, so the values can be re-scaled directly from the implementation. Another key concern addressed in all the following regressions is that because the fixed effects are estimated values they might include sampling error, which could overestimate the variance explained by each individual component and affect inference. In appendix B.5 we empirically verify that this error is relatively constant over time (and cross sectionally). On top of this, standard errors are calculated by Efron bootstrap, which leads to a conservative inference (Hahn and Liao, 2021).

5.1.1 Firm Pay Premium

Understanding the factors that determine why different firms pay different wages is essential for understanding wage dynamics. Table 4 reports the resulting coefficients from projecting the estimated firm pay premium into the logarithm of average firm size, average value added per worker, average product market concentration, average labor market concentration measures, and average (skill) knowledge composition. Across subperiods, the predictive power of our model seems to be relatively constant at around 30 percent. The results in the table control for sorting into jobs and job titles, and several aspects deserve special attention: workers who work in firms that are larger, more productive, operate in less competitive environments, and in which the average job skill requirements are higher receive significantly higher wages. The occupational structure of the firm has the largest impact. On the other hand, workers who work in monopsonic firms and firms with a high share of minimum wage workers are expected to be paid less.

A striking finding in table 4 is that the coefficients decrease (in absolute value) from the first to the third subperiod. Alvarez et al. (2018) using this finding for value added per worker, suggest that the relationship between return to characteristics and the firm effects flattens over time, which is consistent with the decline of the passthrough from characteristics to pay (wages). We formally test the decline in passtrough below, but it is worth noting this result upfront. [12]

5.1.2 Job Title Pay Premium

Different types of jobs have different compensation schemes, depending on the occupation, sector, skills and tasks performed, and institutional settings that might regulate the profession. Such dynamics are captured by job titles in specification 1. We regress the estimated job title fixed effects on the selected firm characteristics, to understand why different jobs pay differently. The results are presented in table 5.

[11] See https://github.com/FixedEffects/FixedEffectModels.jl

[12] To compare our findings with previous literature, we evaluate the decline in the coefficient of value added per worker graphically. The results are shown in figure A2 in the appendix. We have verified that this decline in the passthrough is not the result of changes in the firm size distribution over time. When we evaluate this relationship conditional on firm size (measured by the average number of workers in each sub-period), we verify that the decline in value added passthrough is common for all the firm size groups.

Table 4: Projection of Covariates into Firm Fixed Effects (All Periods)

	$\hat{\psi}_j$ - Firm fixed-effects			
Sample	2005-2009	2010-2014	2015-2019	2005-2019
log(Firms size)	0.028***	0.020***	0.018***	0.027***
	(0.000)	(0.000)	(0.000)	(0.000)
log(Value added) p.w.	0.012***	0.011***	0.009***	0.015***
	(0.000)	(0.000)	(0.000)	(0.000)
Product concentration	0.089***	0.081***	0.039***	0.087***
	(0.001)	(0.001)	(0.001)	(0.001)
Labor Market concentration	-0.010***	-0.028***	-0.002***	-0.038***
	(0.001)	(0.001)	(0.001)	(0.000)
Share of MW workers	-0.289***	-0.265***	-0.280***	-0.265***
	(0.000)	(0.000)	(0.000)	(0.000)
Workforce composition	0.401***	0.281***	0.189***	0.375***
	(0.002)	(0.002)	(0.001)	(0.001)
N	6,779,289	6,301,025	6,595,153	22,192,961
R^2	0.340	0.311	0.348	0.428
Significance:				*p<0.1; **p<0.05; ***p<0.01

Sources: *Quadros de Pessoal*, $2005-19$.

Note: The table displays the coefficients obtained when projecting the estimated firm pay premium component into the covariates. All the estimations presented in the table control for sector and region. Standard errors are reported in parentheses. The standard errors are calculated by bootstrap, using 500 repetitions.

Some aspects stand out from the results. First, the explanatory power of the firm variables in the firm job-title fixed effects is lower compared with the size of the resulting coefficients when we projecting the firm pay premium (see table 4). Second, it also seems that in this case the coefficients shrink over time, which suggests that there is a decline in the passtrough. As in the previous case, we quantify this change in the next subsection. Nevertheless the signs of the coefficients coincide for all the variables except the market concentration variables: the job title pay premium increases in jobs that are in concentrated labor markets, while it decreases in markets that have more competition. This could be driven by the workers' ability to organize themselves, and use the institutional tools that the collective agreement provides to increase their negotiation power.

Table 5: Projection of the Covariates into Job-Title Fixed Effects (All Periods)

	$\hat{\phi}_j$ - Job-title fixed-effects			
Sample	2005-2009	2010-2014	2015-2019	2005-2019
log(Firms size)	0.004***	0.002***	0.000***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
log(Value added) p.w.	0.001***	0.001***	0.002***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Product concentration	-0.037***	0.021***	0.011***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Labor Market concentration	0.057***	0.030***	0.007***	0.040***
	(0.001)	(0.001)	(0.000)	(0.000)
Share of MW workers	-0.023***	-0.013***	-0.018***	-0.031***
	(0.000)	(0.000)	(0.000)	(0.000)
Workforce composition	0.355***	0.229***	0.251***	0.455***
	(0.001)	(0.001)	(0.001)	(0.000)
N	6,779,289	6,301,025	6,595,153	22,192,961
R^2	0.276	0.234	0.224	0.297

Significance: *p<0.1; **p<0.05; ***p<0.01

Sources: *Quadros de Pessoal*, 2005 − 19.
Note: The table displays the coefficients obtained when projecting the covariates into the estimated job-title pay premium component. All the estimations presented in the table control for sector and region. Standard errors are reported in parenthesis. The standard errors are calculated by bootstrap, using 500 repetitions.

Moreover, table 5 provides evidence that if the workforce composition in the firm is highly skilled, it increases the job title pay premium. It is then relevant that the workforce structure is associated with the task intensity and degree of technology used by the firm. It is interesting to note how this factor passes to wages through the types of jobs held by the workers. Moreover, as almost all the variable coefficients are positive, we might infer that there is positive sorting between good firms and good job titles. The variance decomposition in table 2 shows that this is indeed the case.

5.2 Second Moment of the Fixed Effect Distribution

Section 5.1 focused on the mean of the distribution of fixed effects. In this section, we are interested in the degree of dispersion of the distribution, identified by the variance —the second moment of the distribution. To see the changes in the variance and isolate the role of firm fundamentals in the passthrough or change in structure, it is common in the literature to apply a two step procedure.[13] The result computes the variance of the fixed effects explained by observable characteristics. Instead of relying on a post-estimation procedure, we propose to project the covariates of interest into the non-parametric variance counterpart, which is also known as the RIF —recentered influence function —(Firpo et al., 2009, 2018). This methodology allows us to estimate directly how the observables affect the dispersion of the variable of interest, which in this case, is how firm characteristics affect the firm pay premium variance, job title pay premium variance, and their covariance. More importantly, we can quantify the impact of each firm observable and divide the effect into the passthrough and structural effects. In this way, we can identify the drivers of the wage inequality dynamics. To calculate the non-parametric version of the variance we use the following:

$$\text{RIF}_{(\sigma_x^2),j} = \left(x_j - \int x dF(x)\right)^2$$

For the covariance we use the following,

$$\text{RIF}_{(\sigma_{x,y}^2),j} = \left(x_j - \int x dF(x)\right)\left(y_j - \int y dF(y)\right)$$

Intuitively $\text{RIF}_{(\sigma_x^2),j}$ and $\text{RIF}_{(\sigma_{x,y}^2),j}$ calculate the marginal contribution of observation j to each statistic. Thus, and following Firpo et al. (2009), for each subperiod P, we can estimate the change in the dispersion by a change in the firm characteristics.

$$\sigma_{\hat{y}}^{2P} = \text{RIF}_{(\sigma_{\hat{y}}^2),j}^{P} = \bar{\mathbf{X}}_j^P \beta_j^P + s_j^P + r_j^P + v_j^P \tag{6}$$

where $\text{RIF}_{(\sigma_{\hat{y}}^2),j}$ stands for the individual contribution to the variance of the estimated firm or job-title fixed-effects in each of the sub-periods. $\bar{\mathbf{X}}_j$ is a matrix of the average firm characteristics for firm j in each subperiod, and v_j stands for an i.i.d. error term. As before, the matrix $\bar{\mathbf{X}}_j$ includes the firm characteristics that are relevant for our exercise, and our specification also accounts for firm sector and region fixed effects. Subection 5.2.1 reports the findings on the dispersion; the tables reporting the firm fixed-effects (table A2), job title fixed-effects (table A3) and their covariance (table A4) are in the appendix. Subsection 5.2.2 decomposes the variance over time and divides the contributions of individual, single observables into pass-through and structural effects.

5.2.1 Fixed-Effects Dispersion

Table A2 presents the results of regressing firm characteristics on the contribution of firm pay premium variance. The regressions on the dispersion of the firm pay premium suggest that as firm size increases, the pay premium decreases, that is, there is less

[13] To calculate the variance, the estimated coefficients $\mathbf{b}$ are multiplied by the variance of the design matrix $\mathbf{X}$: $\text{var}(\hat{y}) = \boldsymbol{b}'\text{var}(\mathbf{X})\boldsymbol{b}$. To compare the change in pass-through in time the design matrix is held constant, and to see the change in observable characteristics the $\mathbf{b}$ is held constant. The counterfactual exercise allows to determine the source of the decline.

heterogeneity in the firm pay premium between larger firms than between smaller firms. Instead, there is increased heterogeneity in all the other variables. In firms that have larger productivity per worker, compete in less competitive markets, have more local labor market concentration, have a larger share of workers earning the minimum wage, and have a more specialized workforce, there is an increase in the dispersion of firm pay premium. As for the results for the first moment of the distribution, the coefficients decrease (in absolute value) from the first to the third subperiod. Such result implies that in the hypothetical case in which the structure of the covariates does not change across periods, the passthrough of the firm characteristics to the firm pay premium falls from the first to the third sub-period.

The correlation of the covariates with the job-title pay premium variance are all small and positive. The exception is product concentration, which is not significant. Table A3 provides support for how the firm characteristics increase dispersion. Although the resulting coefficients are small, the proportion of the variance explained by the model is between 8 and 19 percent. Unlike the firm pay premium, not all the coefficients decrease between the first and third periods, so the preliminary analysis of the passtrough is inconclusive. In this case, product concentration flips sign and increases, and size does not change.

One of the advantages of using the non parametric version of the covariance is that we can calculate the effect of the firm characteristics directly in the covariance. So we can explain what are the factors that increase or decrease the assortativity between good firms and good types of jobs. Table A4 presents the results of the projection of the selected covariates into the covariance. The proportion of the variance explained by the firm characteristics in this case is larger than that in the two previous regressions, and the model explains between 12 and 19 percent of the total variation. The magnitudes of the coefficients, as in the job title pay premium are smaller than in the firm pay premium. In this case, we cannot perform a preliminary exercise to determine the importance and direction of the passthrough.

One of the purposes of this section is to quantify the importance and identify the firm characteristics that drove the decrease in the firm pay premium, job title pay premium and their covariance. Even if the results from the second moment elucidate the variables that are more important in increasing pay for different pay premium dispersions, they are not sufficient to assess what was the principal driver of the reduction in inequality.

5.2.2 Passthrough and Composition

To decompose the change in variance throughout the period in consideration, we use an Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973; Kitagawa, 1955; Card et al., 2016; Firpo et al., 2018). The decomposition allows us to identify how much of the reduction over time in the fixed effects was due to changes in the distribution of the covariates, and the amount of the decrease that was due to a reduction of the passthrough from the firm characteristics to the pay premiums. The following equation is used for the decomposition:

$$\hat{\Delta}^{\sigma^2} = \underbrace{(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0)'\hat{\beta}_0}_{\text{Composition Effects}} + \underbrace{\bar{\mathbf{X}}_1'(\hat{\beta}_1 - \hat{\beta}_0)}_{\text{Pass-through}} \tag{7}$$

where $\hat{\Delta}^{\sigma^2}$ is the total change in the variance, where the composition effects are calculated by fixing the returns from the characteristics and changing the distribution of the design matrix over time. The passtrough is calculated by fixing the design matrix and calculating the changes in the return from the characteristics.

The results of the decomposition are presented in table 6. To calculate the standard errors, we bootstrap the whole procedure using 500 replications. We follow Hahn and Liao (2021) to perform the bootstrapping. The top section of the table shows the aggregate decomposition, for each of the fixed effects that lead to the decrease in inequality. The first two columns assess the reduction in the variance of the firm pay premium. Columns 3 and 4, decompose the change in the job title pay premium and the last two columns decompose the decrease in sorting between firms and job titles. The bottom section of the table provides the details of the decomposition and how each firm covariate contributes to the reduction in inequality in each component. For all the fixed effects, the driver of the decrease in the pay premium is due to a decrease in the passthrough. In all cases, the passthrough overcompensates the increase in variance of the distribution of the characteristics.

Focusing on the passtrough, from the last to the first subperiod, all the covariates except the logarithm of size, the intercept, the region, and the sector, explain the compression in the dispersion of the firm pay premium. In particular, both firm performance proxies are drivers of the fall in inequality dynamics. The first decomposition considers the change that occurred before the financial crisis, which could be the reason why the signs of some of the variables change, specifically for value added and firm size.

In the case of the dispersion in the job title pay premium, all the covariates except the intercept, product market concentration, and region contribute to the decrease in the passthrough. The four most important components that drive the fall in the passthrough are value added, labor market concentration, occupational structure of the firm, and sector. In this component, the sector is the main driver of the fall of the passthrough. When we consider the change in the distribution of firm observable characteristics, its change contributes to an increase the dispersion. However, this contribution is rather small and the passthrough dominates the overall effect.

When we consider the assortativity between the firm and job title components, its decrease was due to the role of firm performance (both value added and occupational structure) along with the contribution of sectors. Looking at structural changes over subperiods, the contribution of the distribution of characteristics is small and increases the dispersion. In this case, the variable that drives the fall in assortativity is the firm sector.

Table 6: Oaxaca-Blinder decomposition for changes in the firm fixed effects variance (RIF)

	Firm Fixed Effects		Job Title Fixed Effects		Cov(JT,F) Fixed Effects	
	$\Delta\sigma^2_{\psi(1.2)} * 100$	$\Delta\sigma^2_{\psi(1.3)} * 100$	$\Delta\sigma^2_{\phi(1.2)} * 100$	$\Delta\sigma^2_{\phi(1.3)} * 100$	$\Delta\sigma^2_{(\psi,\phi)(1.2)} * 100$	$\Delta\sigma^2_{(\psi,\phi)(1.3)} * 100$
Total effects	-0.869	-2.379	-0.244	-0.529	-0.109	-0.296
Pass-through	-1.097	-2.807	-0.444	-0.799	-0.193	-0.435
Composition effects	0.228	0.427	0.201	0.270	0.084	0.139
Total effects						
log(Firms size)	-0.342***	0.475***	0.098***	-0.091***	-0.015***	-0.023***
Excess log(Value added) p.w.	0.553***	-1.203***	-0.867***	-0.577***	-0.508***	-0.394***
Product concentration	-0.163***	-0.165***	0.238***	0.378***	0.095***	0.121***
Labor Market concentration	-0.096***	-0.087***	-0.318***	-0.562***	-0.111***	-0.214***
Share of MW workers	-0.423***	-0.443***	-0.043***	-0.064***	-0.107***	-0.122***
Workforce composition	-2.235***	-3.471***	-0.988***	-0.198***	-0.414***	-0.306***
Region	0.006	-0.031**	-0.07***	0.062***	-0.008***	0.029***
Sector	1.758***	0.607	-0.94***	-1.942***	-1.213***	-1.255***
Pass-through						
Intercept	0.073	1.938***	2.646***	2.466***	2.172***	1.869***
log(Firms size)	-0.241***	0.637***	0.066***	-0.141***	-0.028***	-0.044***
Excess log(Value added) p.w.	0.573***	-1.214***	-0.855***	-0.583***	-0.501***	-0.398***
Product concentration	-0.17***	-0.177***	0.239***	0.378***	0.096***	0.122***
Labor Market concentration	-0.101***	-0.096***	-0.35***	-0.614***	-0.123***	-0.234***
Share of MW workers	-0.582***	-0.858***	-0.061***	-0.111***	-0.141***	-0.213***
Workforce composition	-2.562***	-3.793***	-1.107***	-0.316***	-0.472***	-0.363***
Region	0.017	0.003	-0.065***	0.076***	-0.006**	0.035***
Sector	1.896***	0.754*	-0.958***	-1.954***	-1.189***	-1.21***
Composition effects						
log(Firms size)	-0.101***	-0.163***	0.031***	0.05***	0.013***	0.022***
Excess log(Value added) p.w.	-0.021***	0.011***	-0.012***	0.006***	-0.007***	0.004***
Product concentration	0.007***	0.012***	0*	0*	-0.001***	-0.002***
Labor Market concentration	0.006***	0.009***	0.032***	0.052***	0.012***	0.019***
Share of MW workers	0.159***	0.416***	0.018***	0.047***	0.035***	0.091***
Workforce composition	0.327***	0.322***	0.119***	0.118***	0.057***	0.057***
Region	-0.011***	-0.033***	-0.006***	-0.014***	-0.002***	-0.006***
Sector	-0.138***	-0.147***	0.018***	0.012***	-0.024***	-0.045***

Significance: *p<0.1; **p<0.05; ***p<0.01

6 Drivers of Within Firm Inequality

In addition to investigating the sources of between firm dynamics, we analyze the drivers of within-firm inequality. In this section, we leverage a unique annual survey on firms' adoption of information and communication technologies (IUTICE) to investigate the channels that lie behind the dynamics in *within-firm* inequality. What makes this survey unique is both its ability to be merged with our administrative labor data and the fact that it contains variables that can be seen as close proxies for technology (use of computers, adoption of some form of big data technologies). Guided by the conceptual framework laid down in section 2, technology shows up as a natural candidate. Upfront, the direction of the effect of technology on within-firm inequality is not altogether clear. On the one hand, it might be that increments in the level of technological complexity of a firm's structure act to increase within-firm wage dispersion via the need for higher workforce specialization within the firm to operate such technology. On the other hand, it is possible that technology adoption occurs concurrently with increased specialization of services and outsourcing so that firms' pay policy becomes more homogeneous, hence leading to less within-firm inequality.

6.1 Matching

To determine which direction stands out from the data, we contrast within-firm inequality for firms that adopted big data technologies in 2016 and 2018 with that of otherwise similar firms that did not. Big data-specific questions show up exclusively on the survey in 2016 and 2018, thereby binding our choice of years. Importantly, we classify a given firm j as a big data adopter in a given year t if the firm answered affirmatively to using big data in *at least* one big data component question in the IUTICE survey. Moreover, we take the variance of within-firm (log) hourly wages as our measure for within firm inequality and use propensity score matching (PSM) to match big data adopters to otherwise similar counterparts on the basis of (log) firm sales, lagged firm sales (first- and second-order lags), (log) firm size, lagged firm size (first- and second-order lags), region, industry, and firm quality. In the process of matching, we restrict our pool of counterfactual donors to firms that were never treated (neither in 2016 nor in 2018). This avoids selecting a late treated firm (that is, one being treated exclusively in 2018) as a control firm in the early period (2016). We thus end up with a group of treated firms (big data adopters) and a group of placebo firms (big data non-adopters) over two repeated cross-sections (2016 and 2018). Moreover, we allow for replacement in the process of matching so that in principle the same firm may act as a counterfactual for more than a single treated firm.

6.2 Balance Checks

Critically, however, the validity of our inference depends on the extent to which the two groups are indeed balanced in terms of their characteristics. We verify that this is the case using two different approaches. First, figures (A3), (A4), and (A5) show how propensity score matching makes the otherwise dissimilar distributions of firm quality, firm sales, and firm size, respectively, similar across treated and non-treated units for both years. Second, tables (7) and (A5) present balance checks supporting the claim that the treated and non-treated units are similar in terms of their observables.

Table 7: Balance Check for Matching: Implementation of Big Data and Control Group

	Control Mean	Control SD	Treated Mean	Treated SD	Abs. Std. Diff
$\hat{\psi}$	-0.0221	0.215	-0.0631	0.245	0.18
N_{t-1}	5.024	1.464	4.856	1.682	0.10
S_{t-1}	17.08	1.865	16.89	2.238	0.096
S_{t-2}	17.09	1.832	17.35	1.796	0.14
N_{t-2}	5.033	1.429	5.142	1.534	0.07
N_t	413.8	1245	369.9	1313	0.03
S_t	7.970e+07	2.251e+08	9.400e+07	3.830e+08	0.05

Sources: *IUTICE*, 2016 and 2018.
Note: This table displays balance checks for the control and treatment groups in terms of employment, estimated firm fixed effects, and sales. The last column present the absolute (standardized) mean difference for each variable.

6.3 Empirical Model and Results

Let $T_{jt} = \{0, 1\}$ designate an indicator variable for whether or not firm j adopted big data technologies in year t, with $t = \{2016, 2018\}$. We are primarily interested in estimating the following extended model:

$$\sigma^2_{jt} = \alpha_o + \delta_t + \alpha_1 T_{jt} + \Gamma X_{jt} + \eta_{jt} \tag{8}$$

where σ^2_{jt} is the within-firm j variance of wages in year t and X_{jt} is a vector of firm characteristics including firm j estimated firm fixed effects in the third subperiod (acting as a proxy for firm quality), the firm-specific value added Herfindhal Hirschmann Index, the share of workers earning the minimum wage in firm j and a workforce composition index. In some of the specifications, this vector also includes region and industry fixed effects. δ_t stands for a drift term. Finally, η_{jt} is an orthogonal error term such that

$$\mathbb{E}[\eta_{jt}|T_{jt}, X_{jt}] = 0 \tag{9}$$

We estimate model (8) by pooled ordinary least squares (OLS) using robust standard errors. Our primary coefficient of interest is $\hat{\alpha}_1$, which captures the effect of big data adoption on within-firm inequality. Furthermore, we weight all the regressions by firm size. Table (8) presents the main results of estimating model (8). We find a positive and sizable effect of big data adoption on within firm inequality. The magnitude of this effect can be better appreciated by looking at the median and mean of the variance of log hourly wages for the entire sample of firms considered for the regressions: $Me = 0.13$ and $\bar{x} = 0.16$. This suggests that the adoption of big data increases within firm inequality by as much as 37 percent of the median of the variable's distribution. The direction and magnitude of coefficient $\hat{\alpha}_1$ are highly robust to the addition of the extra covariates in X_{jt}.

Table 8: Treatment Effect of Big Data Adoption

	Dependent variable:							
	Variance of Within Firm Wages							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Group	0.050***	0.043***	0.043***	0.044***	0.045***	0.045***	0.026***	0.027***
	(0.013)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.008)	(0.008)
HH Index				-0.007	-0.008	-0.008	0.011	0.012
				(0.023)	(0.020)	(0.020)	(0.018)	(0.018)
Estimated Firm FE					0.227***	0.225***	0.238***	0.240***
					(0.033)	(0.040)	(0.037)	(0.037)
Share of Min. Wage Workers						-0.002	0.115***	0.120***
						(0.021)	(0.027)	(0.026)
Workforce composition							0.356***	0.343***
							(0.043)	(0.044)
Total Factor Productivity								0.002
								(0.002)
Constant	0.110***	0.125***	0.129***	0.131***	0.144***	0.144***	0.028	0.026
	(0.010)	(0.012)	(0.014)	(0.016)	(0.015)	(0.015)	(0.021)	(0.021)
Region FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sectoral FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,493	2,493	2,493	2,475	2,471	2,471	2,471	2,468
R^2	0.068	0.165	0.166	0.174	0.279	0.279	0.357	0.359
Adjusted R^2	0.068	0.155	0.156	0.164	0.270	0.269	0.349	0.350

Sources: *IUTICE*, 2016 and 2018.
Note: This table displays the treatment effect of big data adoption on within-firm wage dispersion as measured by the within-firm variance of log hourly wages. The top row coefficients capture the effect of big data adoption on within-firm variance of wages, for otherwise similar firms. Significance at the 1 percent level is indicated with three stars, significance at the 5 percent level is indicated with two stars.

7 Conclusion

Using a rich combination of linked employer-employee administrative data and survey data on technology adoption for Portugal, we examined the channels through which firms affect earnings inequality dynamics.

In the absence of firm-specific effects, the worker and sorting components would have driven up inequality, in line with what happened in other advanced economies. The firm-specific pay policy dispersion, the job title pay premium, and their covariance are responsible for the sharp decline in wage inequality. In our conceptual framework, firm pay concentration depends not just on the distribution of firm characteristics (*a composition effect*), but also on a scaling term that dictates the extent to which the dispersion of those characteristics effectively translates into dispersion of the firm pay premium (*a passthrough effect*). Using an Oaxaca-Blinder decomposition, we found that the reduction in the premiums of firm characteristics, rather than changes in their distribution, drove the decline in earnings inequality.

We quantify the contribution of firm size, performance, labor market concentration, and the share of workers earning the minimum wage for changes in firm pay premium dispersion. We found that value added and the firm workforce composition contributed positively to the dispersion of firm fixed effects. Moreover, we found that these two factors were the main contributors to the fall in the firm pay premium dispersion, and this effect came from a fall in the passthrough from firm characteristics to pay.

We also found that for changes in the within-firm component of wage inequality, technology adoption played a key role. Leveraging a survey on the adoption and diffusion of information technologies, we estimated the causal effect of big data adoption on the dispersion of within-firm wages. Our within-firm analysis showed that big data adoption by firms, increases within-firm inequality by as much as 37 percent of the median of within-firm inequality. This result is consistent with empirical evidence showing that automation may increase inequality through skill-biased adoption or declining wages for workers who are specialized in routine tasks in sectors witnessing steep automation (Card and DiNardo, 2002; Acemoglu and Restrepo, 2019).

Our findings suggest that policies that limit product market concentration or foster technology adoption for low-technology firms may be effective in addressing inequality.

References

Abowd, John M. and Francis Kramarz, "Chapter 40 The analysis of labor markets using matched employer-employee data," *Handbook of Labor Economics*, January 1999, *3*, 2629–2710. Publisher: Elsevier. (Cited on page(s) 6)

_ , _ , and David N. Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333. (Cited on page(s) 3, 7, 12, 51)

_ , Robert H. Creecy, and Francis Kramarz, "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Technical Report 2002-06, Center for Economic Studies, U.S. Census Bureau March 2002. Publication Title: Longitudinal Employer-Household Dynamics Technical Papers. (Cited on page(s) 9)

Acemoglu, Daron and Pascual Restrepo, "Automation and new tasks: how technology displaces and reinstates labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30. (Cited on page(s) 5, 6, 15, 27)

_ and _ , "Robots and jobs: Evidence from US labor markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244. Publisher: The University of Chicago Press Chicago, IL. (Cited on page(s) 15)

_ and _ , "Demographics and Automation," *The Review of Economic Studies*, June 2021, (rdab031). (Cited on page(s) 15)

Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser, "Firms and the Decline in Earnings Inequality in Brazil," *American Economic Journal: Macroeconomics*, January 2018, *10* (1), 149–189. (Cited on page(s) 2, 3, 4, 6, 7, 12, 14, 16, 17, 41, 44, 46, 51)

Andrews, M. J., L. Gill, T. Schank, and R. Upward, "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2008, *171* (3), 673–697. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-985X.2007.00533.x. (Cited on page(s) 3, 6, 48)

Atkinson, Anthony B., "More on the measurement of inequality," *Journal of Economic Inequality*, 2008, *6* (3), 277. Publisher: Springer Nature BV. (Cited on page(s) 43)

Autor, David, Claudia Goldin, and Lawrence F. Katz, "Extending the Race between Education and Technology," *AEA Papers and Proceedings*, May 2020, *110*, 347–351. (Cited on page(s) 11)

Autor, David H. and Michael J. Handel, "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of labor Economics*, 2013, *31* (S1), S59–S96. Publisher: University of Chicago Press Chicago, IL. (Cited on page(s) 6)

Azar, José, Steven Berry, and Ioana Elena Marinescu, "Estimating labor market power," *Available at SSRN 3456277*, 2019. (Cited on page(s) 15)

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman, "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States," *Journal of Labor Economics*, April 2016, *34* (S2), S67–S97. Publisher: The University of Chicago Press. (Cited on page(s) 6)

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources*, 1973, *8* (4), 436–455. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System]. (Cited on page(s) 5, 21)

Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter, "The Disappearing Large-Firm Wage Premium," *AEA Papers and Proceedings*, May 2018, *108*, 317–322. (Cited on page(s) 6, 14, 46)

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 2019, *87* (3), 699–739. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15722. (Cited on page(s) 2, 3, 6, 9, 47)

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 2006, *74* (2), 323–364. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2006.00665.x. (Cited on page(s) 7)

Card, David, Ana Rute Cardoso, and Patrick Kline, "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women," *The Quarterly Journal of Economics*, May 2016, *131* (2), 633–686. (Cited on page(s) 9, 15, 21)

_ and _ , "Wage Flexibility Under Sectoral Bargaining," SSRN Scholarly Paper ID 3828330, Social Science Research Network, Rochester, NY April 2021. (Cited on page(s) 9, 16)

_ and John E. DiNardo, "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, October 2002, *20* (4), 733–783. Publisher: The University of Chicago Press. (Cited on page(s) 27)

_ , Jörg Heining, and Patrick Kline, "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 2013, *128* (3), 967–1015. (Cited on page(s) 2, 4, 6, 7, 12, 41, 47)

Cardoso, Ana Rute, Paulo Guimarães, and Pedro Portugal, "What drives the gender wage gap? A look at the role of firm and job-title heterogeneity," *Oxford Economic Papers*, 2016, *68* (2), 506–524. Publisher: Oxford University Press. (Cited on page(s) 6)

_ , _ , _ , and Pedro S. Raposo, "The sources of the gender wage gap," *Banco de Portugal Economic Studies*, 2016, *2016*, 2. (Cited on page(s) 6, 47)

Carneiro, Anabela, Paulo Guimarães, and Pedro Portugal, "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity," *American Economic Journal: Macroeconomics*, April 2012, *4* (2), 133–152. (Cited on page(s) 6, 9)

—, **Pedro Portugal, Pedro Raposo, and Paulo M. M. Rodrigues**, "The persistence of wages," *Journal of Econometrics*, January 2022. (Cited on page(s) 9)

Devicienti, Francesco, Bernardo Fanfani, and Agata Maida, "Collective bargaining and the evolution of wage inequality in Italy," *British Journal of Industrial Relations*, 2019, *57* (2), 377–407. Publisher: Wiley Online Library. (Cited on page(s) 16)

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske, "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment," *Journal of Labor Economics*, April 2004, *22* (2), 397–429. Publisher: The University of Chicago Press. (Cited on page(s) 6)

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg, "Revisiting the German Wage Structure," *The Quarterly Journal of Economics*, 2009, *124* (2), 843–881. (Cited on page(s) 6)

Fanfani, Bernardo, "The Employment E ects of Collective Bargaining," Technical Report def095, Università Cattolica del Sacro Cuore, Dipartimenti e Istituti di Scienze Economiche (DISCE) October 2020. Publication Title: DISCE - Working Papers del Dipartimento di Economia e Finanza. (Cited on page(s) 16)

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux, "Unconditional Quantile Regressions," *Econometrica*, 2009, *77* (3), 953–973. Publisher: [Wiley, The Econometric Society]. (Cited on page(s) 4, 12, 20)

Firpo, Sergio P., Nicole M. Fortin, and Thomas Lemieux, "Decomposing Wage Distributions Using Recentered Influence Function Regressions," *Econometrics*, June 2018, *6* (2), 28. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. (Cited on page(s) 4, 12, 20, 21)

Gastwirth, Joseph L., "A general definition of the Lorenz curve," *Econometrica: Journal of the Econometric Society*, 1971, pp. 1037–1039. Publisher: JSTOR. (Cited on page(s) 43, 49)

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card, "Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil," *American Economic Review*, October 2021, *111* (10), 3418–3457. (Cited on page(s) 12)

Goos, Maarten and Alan Manning, "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 2007, *89* (1), 118–133. (Cited on page(s) 6)

Guimaraes, Paulo and Pedro Portugal, "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 2010, *10* (4), 628–649. Publisher: SAGE Publications Sage CA: Los Angeles, CA. (Cited on page(s) 48)

Hahn, Jinyong and Zhipeng Liao, "Bootstrap Standard Error Estimates and Inference," *Econometrica*, July 2021, *Vol. 89* (No. 4), 1963–1977. (Cited on page(s) 17, 22)

Kitagawa, Evelyn M., "Components of a Difference Between Two Rates," *Journal of the American Statistical Association*, 1955, *50* (272), 1168–1194. Publisher: [American Statistical Association, Taylor & Francis, Ltd.]. (Cited on page(s) 5, 21)

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten, "Leave-Out Estimation of Variance Components," *Econometrica*, 2020, *88* (5), 1859–1898. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16410. (Cited on page(s) 3, 6, 9, 47)

Leung, Justin H., "Minimum Wage and Real Wage Inequality: Evidence from Pass-Through to Retail Prices," *The Review of Economics and Statistics*, 2021, *103* (4), 754–769. (Cited on page(s) 16)

Lewis, Jeffrey B. and Drew A. Linzer, "Estimating Regression Models in Which the Dependent Variable Is Based on Estimates," *Political Analysis*, 2005, *13* (4), 345–364. Publisher: [Oxford University Press, Society for Political Methodology]. (Cited on page(s) 52)

Lise, Jeremy and Fabien Postel-Vinay, "Multidimensional skills, sorting, and human capital accumulation," *American Economic Review*, 2020, *110* (8), 2328–76. (Cited on page(s) 54)

Messina, Julian and Joana Silva, "Twenty Years of Wage Inequality in Latin America," *The World Bank Economic Review*, February 2021, *35* (1), 117–147. (Cited on page(s) 2, 6, 7, 44, 45)

Mortensen, Dale T. and Christopher A. Pissarides, "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 1994, *61* (3), 397–415. (Cited on page(s) 7)

Mueller, Holger M., Paige P. Ouimet, and Elena Simintzi, "Wage Inequality and Firm Growth," *American Economic Review*, May 2017, *107* (5), 379–383. (Cited on page(s) 6)

Naidu, Suresh and Eric A. Posner, "Labor Monopsony and the Limits of the Law," *Journal of Human Resources*, June 2021, p. 0219. Publisher: University of Wisconsin Press. (Cited on page(s) 15)

_ , _ , and Glen Weyl, "Antitrust remedies for labor market power," *Harv. L. Rev.*, 2018, *132*, 536. Publisher: HeinOnline. (Cited on page(s) 15)

Oaxaca, Ronald, "Male-female wage differentials in urban labor markets," *International economic review*, 1973, pp. 693–709. Publisher: JSTOR. (Cited on page(s) 5, 21)

Portugal, Pedro and Ana Rute Cardoso, "Disentangling the minimum wage puzzle: an analysis of worker accessions and separations," *Journal of the European Economic Association*, 2006, *4* (5), 988–1013. Publisher: Oxford University Press. (Cited on page(s) 16)

_ , Pedro S. Raposo, and Hugo Reis, "The distribution of wages and wage inequality," *Economic Bulletin and Financial Stability Report Articles and Banco de Portugal Economic Studies*, 2018. Publisher: Banco de Portugal, Economics and Research Department. (Cited on page(s) 4, 6)

Postel–Vinay, Fabien and Jean-Marc Robin, "Equilibrium wage dispersion with worker and employer heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350. (Cited on page(s) 7)

Raposo, Pedro, Pedro Portugal, and Anabela Carneiro, "The Sources of the Wage Losses of Displaced Workers The Role of the Reallocation of Workers into Firms, Matches, and Job Titles," *Journal of Human Resources*, January 2021, *56* (3), 786–820. Publisher: University of Wisconsin Press. (Cited on page(s) 6, 9)

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter, "Firming up inequality," *The Quarterly journal of economics*, 2019, *134* (1), 1–50. Publisher: Oxford University Press. (Cited on page(s) 2, 3, 6, 7, 12, 44, 47, 51)

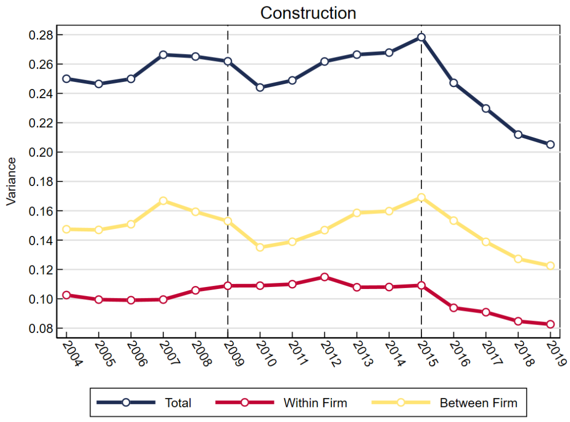
Sorkin, Isaac, "The Role of Firms in Gender Earnings Inequality: Evidence from the United States," *American Economic Review*, May 2017, *107* (5), 384–387. (Cited on page(s) 6)

A Additional Figures and Tables

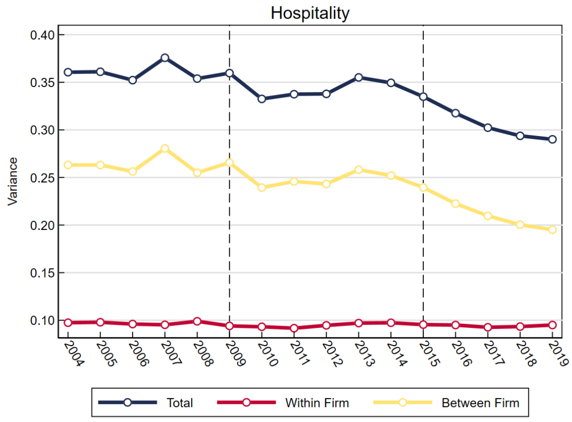
This appendix presents further evidence, and clarifications that are not contained in the body of the paper but might be of interest to the reader.

A.1 Additional Figures and Tables

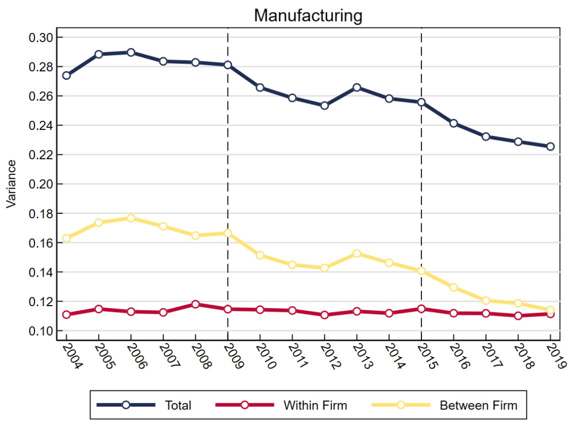
Figure A1: Within and Between Firm Inequality, by Sectors (2005-2019).



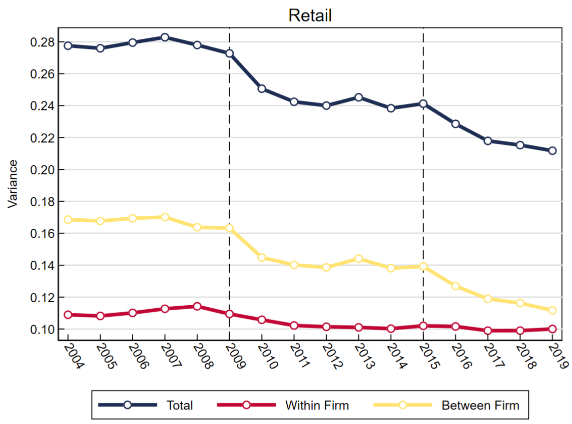
(a)



(b)



(c)



(d)

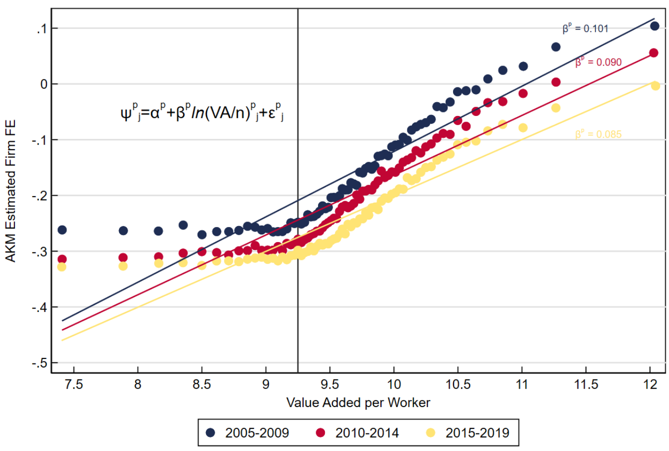
Sources: *Quadros de Pessoal*, $2005-2019$.
Note: Panels (a) - (d) plots the yearly evolution of the variance of hourly wages ("total wage inequality") for the 2005-2019 period, decomposed in a within firm inequality and a between firm inequality components for selected sectors: construction, hospitality, manufacturing and retail.

Table A1: Mobility matrix

Year	Firm Switch	Job Switch	Share of Job switchers	Share of Firm Switchers
2006	0.522	0.311	0.189	0.316
2007	0.487	0.316	0.185	0.285
2008	0.479	0.359	0.209	0.279
2009	0.445	0.278	0.179	0.287
2010	0.291	0.281	0.182	0.188
2011	0.388	0.228	0.138	0.235
2012	0.360	0.972	0.632	0.234
2013	0.362	0.505	0.335	0.241
2014	0.372	0.298	0.209	0.261
2015	0.334	0.314	0.214	0.227
2016	0.426	0.326	0.204	0.267
2017	0.484	0.428	0.261	0.296
2018	0.555	0.467	0.281	0.334
2019	0.624	0.396	0.261	0.411

Sources: *Quadros de Pessoal*, 2005 – 2019.
Note: The table displays the number of workers that switch firm or workers that switch job-title each year, for the period 2006-2019.

Figure A2: Value added per worker and firm pay premium



.

Sources: *Quadros de Pessoal*, 2005 – 2019.
Note: The figure shows the average estimated firm effect in each value added per worker bin.

Table A2: Projection of covariates into firm pay premium variance (all periods)

	RIF$_{(\sigma^2_{\hat{\psi}})}$ - Firm fixed-effects variance			
Sample	2005-2009	2010-2014	2015-2019	2005-2019
log(Firms size)	-0.007***	-0.008***	-0.006***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
log(Value added) p.w.	0.003***	0.003***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Product concentration	0.044***	0.015***	0.014***	0.042***
	(0.001)	(0.001)	(0.001)	(0.000)
Labor Market concentration	0.007***	-0.001	-0.000	0.004***
	(0.001)	(0.001)	(0.000)	(0.000)
Share of MW workers	0.083***	0.046***	0.038***	0.069***
	(0.000)	(0.000)	(0.000)	(0.000)
Workforce composition	0.180***	0.112***	0.079***	0.217***
	(0.001)	(0.001)	(0.001)	(0.001)
N	6,779,289	6,301,025	6,595,153	22,192,961
R^2	0.051	0.081	0.067	0.110

Significance: *p<0.1; **p<0.05; ***p<0.01

Sources: *Quadros de Pessoal*, $2005-2019$.

Note: The table displays the coefficients obtained when projecting the covariates into the estimated firm pay premium variance (RIF$_{(\sigma^2_{\hat{\psi}})}$). All the estimations presented in the table control for sector and region. Standard errors are reported in parenthesis. The standard errors are calculated by bootstrap, using 500 repetitions.

Table A3: Projection of covariates into job-title pay premium variance (all periods)

	RIF$_{(\sigma^2_{\hat{\phi}})}$ - Job-title fixed-effects variance			
Sample	2005-2009	2010-2014	2015-2019	2005-2019
log(Firms size)	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
log(Value added) p.w.	0.002***	0.001***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Product concentration	-0.002	0.039***	0.061***	0.022***
	(0.001)	(0.001)	(0.001)	(0.000)
Labor Market concentration	0.039***	0.011***	-0.009***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)
Share of MW workers	0.009***	0.006***	0.003***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
Workforce composition	0.066***	0.036***	0.057***	0.083***
	(0.001)	(0.001)	(0.001)	(0.000)
N	6,779,289	6,301,025	6,595,153	22,192,961
R^2	0.181	0.190	0.075	0.202
Significance:				*p<0.1; **p<0.05; ***p<0.01

Sources: *Quadros de Pessoal*, $2005-2019$.

Note: The table displays the coefficients obtained when projecting the covariates into the estimated job-title pay premium variance (RIF$_{(\sigma^2_{\hat{\phi}})}$). All the estimations presented in the table control for sector and region. Standard errors are reported in parenthesis. The standard errors are calculated by bootstrap, using 500 repetitions.

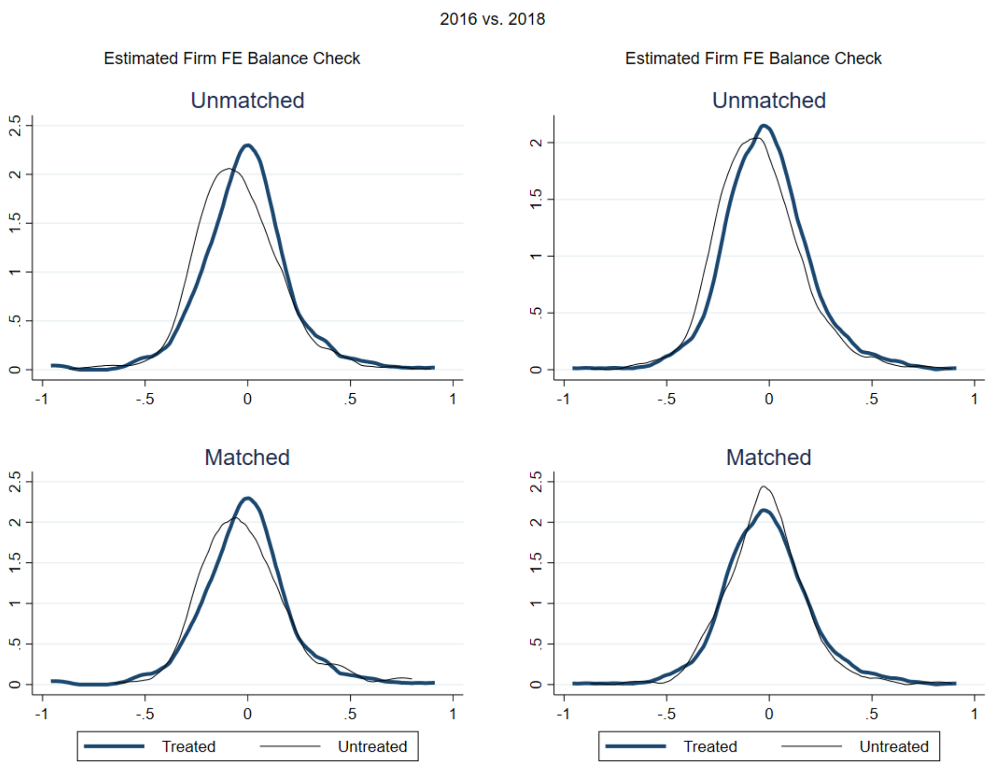
Table A4: Projection of covariates into firm-job-title pay premium covariance (all periods)

	RIF$_{(\sigma^2_{\hat{\psi},\hat{\phi}})}$ - Firm-Job-title fixed-effects covariance			
Sample	2005-2009	2010-2014	2015-2019	2005-2019
log(Firms size)	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Excess log(VA) pw	0.001***	0.001***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Product concentration	-0.006***	0.010***	0.014***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)
Labor Market concentration	0.015***	0.005***	-0.003***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Share of MW workers	0.018***	0.009***	0.007***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)
Workforce composition	0.032***	0.019***	0.022***	0.034***
	(0.000)	(0.000)	(0.000)	(0.000)
N	6,779,289	6,301,025	6,595,153	22,192,961
R^2	0.157	0.204	0.128	0.176
Significance:				*p<0.1; **p<0.05; ***p<0.01

Sources: *Quadros de Pessoal*, $2005-2019$.

Note: The table displays the coefficients obtained when projecting the covariates into the estimated firm-job-title pay premium covariance (RIF$_{(\sigma^2_{\hat{\psi},\hat{\phi}})}$). All the estimations presented in the table control for sector and region. Standard errors are reported in parenthesis. The standard errors are calculated by bootstrap, using 500 repetitions.

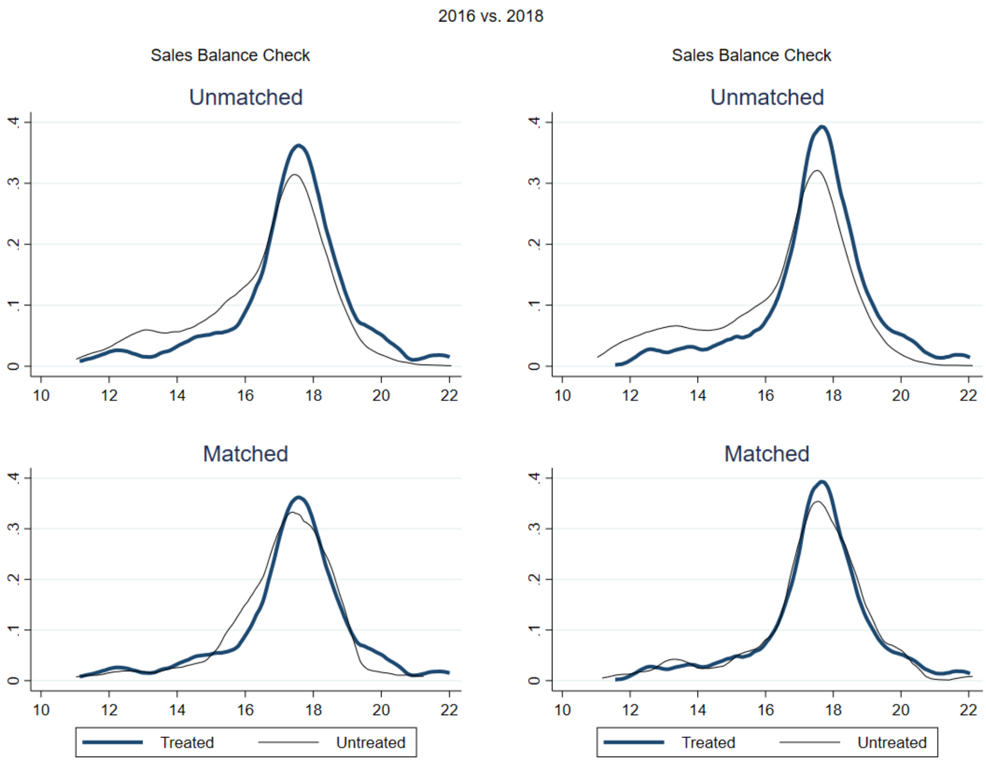
Figure A3: Balance Checks for Control and Treatment: Firm FE



.

Sources: *Quadros de Pessoal*, 2005 – 2019.
Note: This figure shows the distribution of estimated firm fixed effects for treated and untreated units, before and after matching. We show these distributions for the years of 2016 (left-hand side) and 2018 (right-hand side).

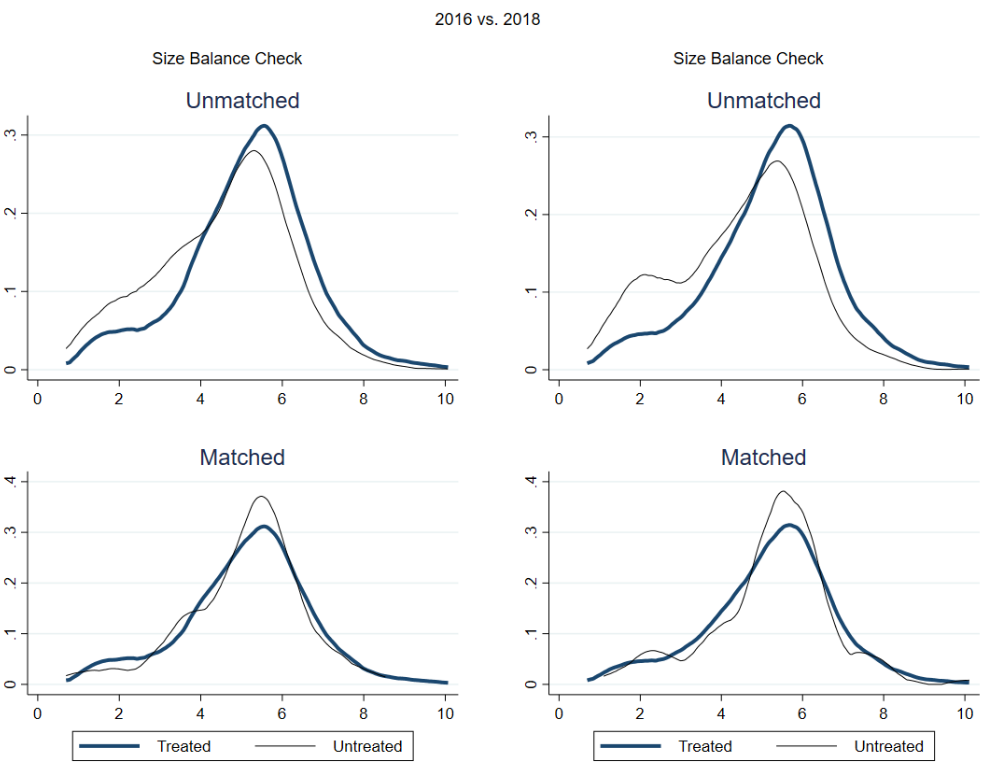
Figure A4: Balance Checks for Control and Treatment: Sales



.

Sources: *Quadros de Pessoal*, $2005 - 2019$.
Note: This figure shows the distribution of estimated firm sales (that is, real revenues) for treated and untreated units, before and after matching. We show these distributions for the years of 2016 (left-hand side) and 2018 (right-hand side).

Figure A5: Balance Checks for Control and Treatment: Size



.

Sources: *Quadros de Pessoal*, 2005 – 2019.
Note: This figure shows the distribution of estimated firm size (that is, employment or number of workers within the firm) for treated and untreated units, before and after matching. We show these distributions for the years of 2016 (left-hand side) and 2018 (right-hand side).

A.2 Exogenous Mobility Assumption

The validity of the AKM results presented in section 2 requires that the match effect is unrelated with firm and worker components, i.e. the error term is indeed strictly exogenous. With this in mind, and in line with the proposal of Card et al. (2013), we look at the average estimated residuals by firm and worker fixed-effects deciles. Figure A6 shows the mean residuals per firm and worker fixed-effect deciles, but also per job and worker fixed effects for selected sub-periods[14]. In support of the AKM hypothesis, the residuals are close to zero, regardless of worker and firm deciles, or worker and job title deciles. Although the residuals do appear larger for lower worker and firm fixed effects combinations, the magnitude is still fairly small – less than 0.04. These results support specification proposed in Equation 1.

Figure A6: AKM Residuals by Firm, Worker and Job-Title Fixed-Effects Deciles.

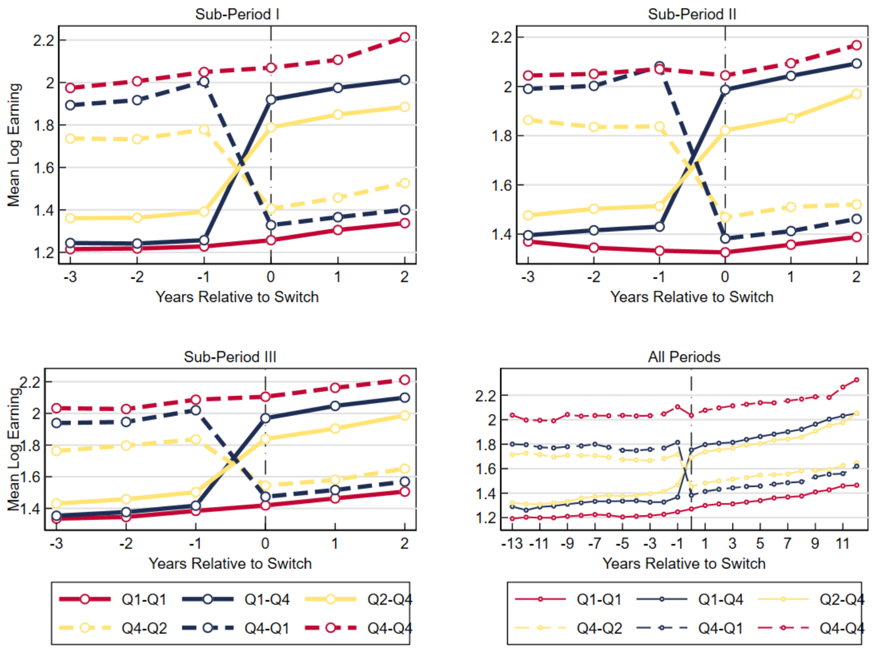
(a) (b)

Sources: *Quadros de Pessoal*, 2005 – 2019.
Note: Panels (a) - (b) the mean residuals per firm and worker fixed-effect deciles, but also per job and worker fixed effects for selected sub-periods.

The second exercise we run to test for exogenous mobility follows Alvarez et al. (2018) and Card et al. (2013). The idea is the following: under AKM assumptions, the gains from those switching, for example, from the first to the forth quartile of firm fixed-effects, should be identical to the losses of those making the opposite switch, for example, from the forth to the first quartile of firm fixed effects. Figure A7 shows the average evolution of earnings of workers changing employers in the first sub-period (2005-2009), the second sub-period (2010-2014), the third sub-period (2015-2019), and the overall time span (2005-2019). Consistent with AKM assumptions, indeed workers moving from the 1^{st} firm quartile to the 4^{th} firm quartile exhibit gains that are similar to the symmetric of those moving from the 4^{th} to the 1^{st} (see solid and dash blue lines in Figure A7. This results hold more generally across sub-periods and across quartile combinations.

[14]The results are similar in all three sub-periods.

Figure A7: Change in Wages of Workers Moving across Firm Quartiles.



Sources: *Quadros de Pessoal*, 2005 – 2019.
Note: The figure shows the average evolution in earnings of workers moving across quartiles of firm estimated AKM fixed-effects. This figure presents for all sub-periods the average evolution in earnings of workers moving across quartiles of firm estimated AKM fixed-effects. To do these computations, we consider only workers that changed their jobs at most once during each sub-period. The solid blue line, for instance, presents the year-on-year average of log hourly wages of workers having moved from the lowest quartile to the highest quartile, where years are normalized in reference to the event. In this exercise, we require that workers are present at least two years prior and two years posterior to the job switch.

B Evolution of Earning Inequality in Portugal

Inequality fell until the Great Recession in European countries, when the trend inverted. Portugal is an exception, where labor income inequality has continued to decline. In this section, we start by presenting the first set of stylized facts on Portugal's rapid decrease in earnings inequality between 2005 and 2019. How is Portugal's labor market different from those of other European countries? The Portuguese labor market is characterized by a high unemployment rate relative to other EU countries and the prevalence of a rigid labor market, which makes it difficult for firms to adjust their labor costs. What are the reforms that might have had an impact on inequality? During the period of analysis, we can clearly observe three significant changes in labor market institutions that may have changed the level of inequality. First, there was a staggering increase in the minimum wage. Second, there was flexibilization of labor market regulations, reducing the high cost of dismissal that characterized the Portuguese economy, and making accessible other work contract arrangements. Third, there was a set of changes to the collective bargaining system. Both flexibilization and modification of the collective agreement legislation would imply an increase in labor earnings inequality.

B.1 Stylized facts

Wage inequality in Portugal declined continuously over the course of the twenty-first century, by a staggering 20 percent. Figure 1 Panel (a), presents the evolution of three inequality indicators, the Gini coefficient of wages, the variance of log wages, and the log ratio of wages at the 90th and 10th percentiles, between 2005 and 2019. The three indicators exhibit the same broad behavior throughout the period and show a pronounced decreasing trend. It would be difficult to determine a priori the direction and the effect along the distribution of such changes in the evolution of wage inequality in Portugal. Instead, we limit ourselves to reporting some stylized facts that guide our analysis.

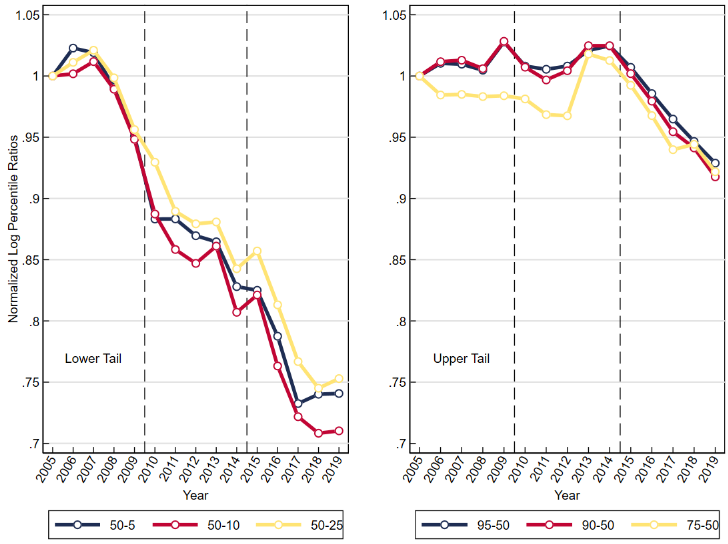
***i)* Heterogeneity in the Change in Inequality across the Labor Income Distribution** Different policies may have different effects along the support of the wage distribution. Even if aggregate inequality was decreasing, different demographic groups along the distribution may have been impacted differently. In what follows, we analyze what happened with (i) the lower tail of the distribution, (ii) the upper tail, and (iii) the distribution as a whole.

Figure A8 presents different measures of inequality over 2005-19. The figure shows that the decrease in inequality in the lower tail of the distribution dominates, compared with the effect in the upper tail. Nevertheless we also observe a decrease in inequality in the upper tail of the distribution, which might indicate that such change happened along the full support of the income distribution. Looking at the normalized log percentile ratios, convergence towards the median of the income distribution occurred at a faster pace for the percentiles below the median, compared with those above the median.

To test this intuition formally, we compare the Lorenz curves of the distributions at the beginning and end of the period considered (Atkinson, 2008; Gastwirth, 1971).[15] Figure A11, in the appendix, supports the claim that inequality *unambiguously* decreased in

[15] A proper derivation of the exercise is presented in appendix B.2.

Figure A8: Wage Inequality Dynamics in Portugal. Upper and Lower Tails (2005-19).



Sources: *Quadros de Pessoal*, $2005-19$.
Note: The figure plots different measures of inequality for the lower and upper tail of the distribution for 2005-19. The inequality measures are normalized to 1 for 2005, and present the evolution of the indicators over time.

Portugal along the support of the distribution. The Lorenz curve for 2019 stochastically dominates the Lorenz curve for 2005, and there are no intersections.

***ii)* Earnings Dispersion Between and Within Firms** For a more in-depth analysis of what might lie behind these trends, we decompose wage inequality into the contributions of within- and between-firm inequality. This provides some preliminary understanding of the role of firm heterogeneity. If all firms paid the same wage to all employees, there would be no within-firm inequality, but not necessarily no wage inequality as firms could still differ in the wages that they pay. Likewise, if all firms had the same distribution of wages there would be no between-firms inequality, but not necessarily no wage inequality as workers within each firm could earn different wages. These are the two extreme cases. With this in mind, we examine which of these factors was more prominent in Portugal between 2005 and 2019, shedding light on whether wage dispersion was mostly driven by systematic differences in pay premiums across firms or differences in pay within each firm. To do so, we decompose the variance of wages into its between and within components (Alvarez et al., 2018; Song et al., 2019; Messina and Silva, 2021),

$$\underbrace{Var(w_t^{i,j,f})}_{Overall\ Inequality} = \underbrace{Var(\overline{w_t^f})}_{Between\ Firm\ Inequality} + \underbrace{\overline{Var(w_t^{i,j,f}|i \in f)}}_{Within\ Firm\ Inequality} \tag{A1}$$

This equation[16] decomposes the yearly overall variance of log real hourly wages into the between-firm component (given by the variance across firm average wages), and the

[16] We present the complete derivation in the appendix B.3.

within-firm component (given by the weighted average of within-firm wage variance, with the weight ω_f being the share of employment in firm f). Throughout the period, between-firm inequality accounted for over 60 percent of total wage inequality, and within-firm inequality accounted for slightly less than 40 percent (see Figure 1, Panel (b)). In the subperiods considered (2005-09; 2010-14; 2015-19) within and between-firm inequality moved broadly in the same direction driving the overall change in inequality. However, the stronger reduction of inequality in the 2010-14 and the 2015-19 subperiods was mostly driven by the reduction in between-firms inequality.

To verify that the observed patterns of between and within inequality are not driven by specific sectors but are representative of the economy as a whole, we further run this equation for four selected sectors: manufacturing, construction, retail, and hospitality (see Figure A1 in the appendix). Our key insight holds regardless of the broad sector being considered.

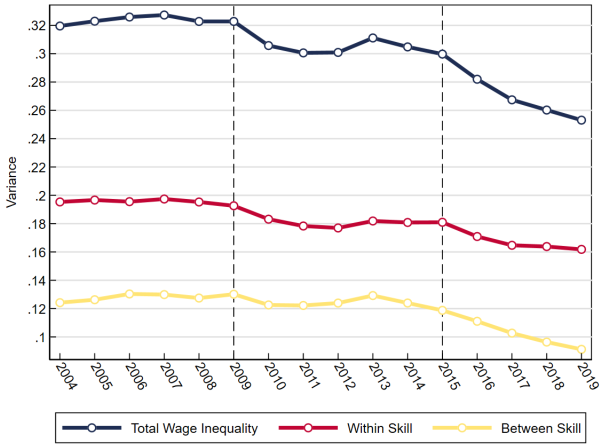
***iii)* Earnings Dispersion Between and Within Skills** The richness of our data also allows us to calculate inequality between different skill groups and assess how this measure has changed over time. This exercise reveals the real prominence of systematic differences in returns to skills across different skill types in determining wage dispersion. To disentangle overall wage inequality, we follow Messina and Silva (2021). We start by running a standard Mincerian equation of the form $w_{it} = \rho_t \mathbf{X}_i + \mu_{it}$, where w_{it} stands for the log hourly wage of worker i in period t. $\mathbf{X}_i$ is a vector of covariates including a categorical educational level variable, tenure by five year bins, and gender (as well as all possible interactions between these). ρ_t is a vector of returns to these covariates, and μ_{it} is an orthogonal error term, referred to as within-skill group wage inequality. Once estimated, we can apply variances to this relation to obtain

$$\underbrace{Var(w_{it})}_{Overall\ Inequality} = \underbrace{Var(\hat{\rho}_t \mathbf{X}_i)}_{Between-Group\ Skill\ Inequality} + \underbrace{Var(\hat{\mu}_{it})}_{Within-Group\ Skill\ Inequality} \tag{A2}$$

where we have used the orthogonality of the error term to impose zero covariance between the residual and the regressors. The variance of wages can thus be decomposed into a between-skill component and a within-skill component. Table A9 below shows the result of this decomposition. In levels, within-skill inequality accounts for the largest share of overall inequality (around 60 percent). In differences, however, between-skill inequality reduction seems to play a role that is roughly as important as within-skill inequality. Over 2005-19, around 50 percent of the reduction in wage inequality is attributed to the reduction in between-skill inequality, against 50 percent explained by within-firm inequality. If we zero in on the inequality reduction witnessed over 2015-19, between-skill inequality reduction accounts for almost 60 percent of the overall reduction in inequality, despite its initially lower level. These findings highlight the importance of considering job titles.

***iv)* Decline in the Large Firm Pay Premium** The role of large firms as providers of better working conditions has been acknowledged in the past: in general, large firms offer better monetary and non-monetary compensation. It is typical that in larger firms,

Figure A9: Between- and Within-Skill Group Inequality (2005 - 19)



Sources: *Quadros de Pessoal*, 2005 – 19.
Note: The figure shows the inequality decomposition of labor income inequality between and within skill groups.

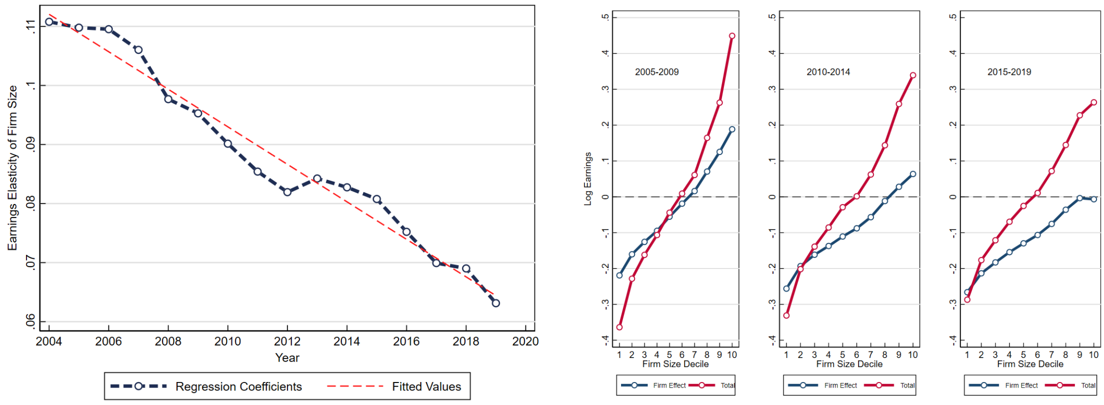
jobs are more stable, there is greater worker satisfaction, and workers earn higher wages. However, there is evidence for the United States that the large-firm wage premium has been shrinking (Bloom et al., 2018). To assess whether this is the case in Portugal, we perform two exercises on the role of large firm size in the wage premium (LFWP). First, we calculate the yearly elasticity of firm size with respect to wages.[17] Second, relying on the estimated firm effects from equation 1, we plot the (de-meaned) average log earnings and average fixed effects for each firm size decile, as in Bloom et al. (2018). This allows us to assess the wage differential between different types of firms over time.

Panel (a) in figure A10 suggests a strongly declining relationship between firm size and wages, dropping from an elasticity of around 11 percent in 2004 to one under 7 percent in 2019, which corresponds to a reduction of approximately 60 percent. This is a sizable drop. Thus, the pay premium that large firms offer appears to have shrunk in absolute terms. This finding is backed by the results presented in panel (b) in figure A10, where we explore this relationship in different periods and the wage differential between large firms and very large firms seems to shrink over time. As large firms have historically paid significantly higher wages, it is important to understand the implications of a fall in the large firm wage premium for changes in inequality.

***v)* Decreased Variance in the Firm Pay premium** The model presented in equation 1 and its variance decomposition in equation 4 consider that changes in variance could be driven by any component on the left-hand side. Worker heterogeneity, firm heterogeneity, job title heterogeneity, or sorting between them could in principle play a role in determining inequality and inequality dynamics. There is evidence that firms might be an important driver of labor income inequality. Alvarez et al. (2018) finds that in Brazil the firm component explains around 39 percent of the decline in wage inequality between 1996 and 2012. In Germany and the United States, the change seems

[17]For each year between 2005 and 2019, we run the following specification: $\log(w_{ij}) = \alpha_o + \alpha_1 \log(N_j) + \epsilon_{ij}$.

Figure A10: Wages and Firm Size in Portugal (2005 - 19)



(a) (b)

Sources: *Quadros de Pessoal*, $2005 - 2019$.
Note: Panel (a) plots the coefficient that results when we project the labor earnings on the size of the firm by year using $\log(w_{ij}) = \alpha_o + \alpha_1 \log(N_j) + \epsilon_{ijt}$. Panel (b) plots the average firm effects and average log earning per firm size decile.

to be driven by better workers sorting into better firms (Card et al., 2013; Song et al., 2019). Which component best explains labor income inequality for the case of Portugal? Which component is driving its decline? In light of our stylized facts, it would seem that firms have driven the decline in wage inequality in Portugal. To provide clear evidence on this matter, we consider the following points of concern.

First we inquire the stability of previous findings in the literature and how robust they are to the inclusion of job title heterogeneity. Occupations and job titles are important determinants in wage formation and therefore in its variance. Although heterogeneity in job titles has been recognized as an important dimension of wage formation (Cardoso et al., 2016b), we lack evidence on its role in the dynamics of wage inequality. The types of skills used in the economy and how they have evolved over time have an impact on how skills and tasks are compensated. Moreover, occupations have different institutional agreements that could vary over time, depending on the performance of certain sectors or workers' strength in negotiation process. Thus, the occupation is a dimension that could explain a large part of wage formation and its variance over time.

Another source of concern is how precise the estimates are. There is bias in the variance from the AKM when it is estimated in the largest connected set, in opposition to the AKM estimated in the *leave-out connected set* sample. The importance of each component in the variance decomposition (equation 4) is sensitive to the sample and the network structure considered. Using Swedish data, Bonhomme et al. (2019)[18] show that the variance of the firm component decreases by almost 40 percent when the variance decomposition is computed in the leave-out sample. Kline et al. (2020), report a 30 percent decrease in the firm component variance when estimated in the Veneto matched employer-employee data[19]. Both results provide evidence of the need to consider such

[18] See table S2 in the appendix of Bonhomme et al. (2019).
[19] The data are generally called Veneto Worker History.

sample to achieve robust calculation of the variance and its dynamics.[20]

One caveat in taking the above concerns into consideration is technical, since the "leave-out connected set" assumes only a two-way fixed effect model of wages. If we consider the job title component, the methodology must be extended to a three-way fixed effects model of wages. Even if the calculation of the largest connected set with multiple high-dimensional fixed effects is known (Guimaraes and Portugal, 2010), the computation of the three-way *leave-one out sample* must be implemented from scratch. Our definition of the leave-one-out sample, extends the KSS methodology to include worker, firm, and job title fixed effects. The three-way leave-one-out sample is constructed under a simple assumption on the network structure: workers are connected to firms, and firms are connected to job titles. We rule out the less restrictive possibility that workers connect directly to both firms and job titles, since that would decrease the probability of finding articulation points in the network. As can be seen in Table 2 the differences between the variances in the two samples are not as large as in previous studies. Including the job title fixed effect allows us to maintain the largest connected sample, for which each component of the variance is close to the variance of each element in the largest connected set.

Table 2 presents the AKM variance decomposition following equation 4 for the three subperiods considered and the whole sample. The inclusion of the job title component clearly gives some insight on the variance decomposition results. The worker effects are the most important source of heterogeneity, followed by the firm effects and sorting between workers and job titles. Even if the job title heterogeneity is not as large as the worker or firm heterogeneity, its size is comparable to that of worker-firm sorting. The results are also consistent for the whole sample. Interestingly, for the whole sample, the worker variance decreases, and worker-firm sorting increases.

Table 3 presents the changes in composition throughout the samples. The last two columns show the changes in the second and third sub-samples with respect to the first period. The last column in Table 3 shows that the negative trend in inequality is totally explained by a decrease in the variance of the firm effect component, the job title, and the covariance of firm and job title effects.

All in all, this section provides a set of stylized facts on the levels and dynamics of Portuguese wage inequality in the twenty-first century. However, these stylized facts raise questions on the drivers of the decrease of inequality. For most of these questions, the role of firms appears to be the mediator for the change happening. The mechanical decompositions hide the fact that different firms may have very different wage profiles due to systematic differences in the workers they hire or the types of jobs. Likewise, similar skills might be rewarded differently across firms. To what extent does the decline in the pay premium that large firms offer drive the inequality dynamics? To what extent is this reduction mediated through firm fixed effects? To shed light on this matter, we investigate the potential channels that affect between- and within-firm inequality.

[20]Andrews et al. (2008) propose another method for correcting the biases in the AKM framework, known also as the trace correction. Yet, the method assumes homoscedasticity. Since we do not assume homoscedasticity, we do not calculate the corrected variance using Andrews et al. (2008)'s correction.

B.2 Changes in Wage inequality along the Labor Income Distribution

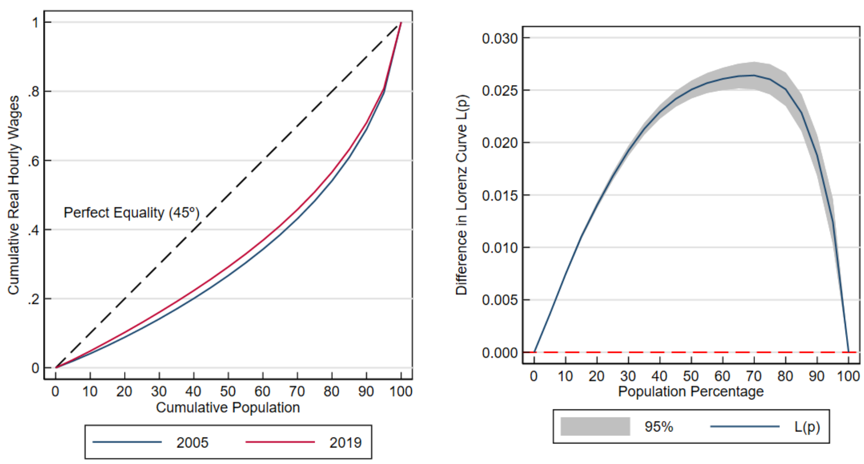
To assess whether inequality *unambiguously* went up or down over the considered period, we evaluate Lorenz's criterion for the log of real hourly wages in Portugal. Specifically, we say that given two distributions X^{2005} and X^{2019}, X^{2019} Lorenz dominates X^{2005} if and only if

$$L_{X^{2019}}(p) \geq L_{X^{2005}}(p) \; \forall p \text{ with } > \text{ for some } p \tag{A3}$$

If this holds, and if Lorenz curves do not cross (since this assures the completeness of the criterion), we can state that X^{2019} is *unambiguously* less unequal than X^{2005}. To perform such exercise empirically, we leverage on Gastworth (1971)'s identity to estimate

$$L_{X^{2019}}(p) - L_{X^{2005}}(p) \Leftrightarrow \frac{1}{\mu^{2019}} \int_o^p Q_X^{2019}(t)dt - \frac{1}{\mu^{2005}} \int_o^p Q_X^{2005}(t)dt \tag{A4}$$

Figure A11: Lorenz Curves for Portugal, in 2005 and 2019.



Sources: *Quadros de Pessoal*, $2005 - 2019$.
Note: The figure shows the Lorenz curve for labor income in 2005 and 2019 (left), and the difference between them. In the left panel the shaded area show the confidence interval at the 95% level for the curve. The confidence interval is calculated by bootstrapping.

Then, one simply has to evaluate whether this differential is positive or negative for $\forall p$. In the expression above, notice how $Q_X(t)$ is the quantile function for the given distribution ("Pen's Parade", the inverse of the cumulative distribution function, CDF), so that estimating $\int_o^p Q_X(t)dt$ boils down to estimating the Generalized Lorenz. When scaled down by the mean of the distribution μ, the Generalized Lorenz becomes the Lorenz curve. The application of this criterion to Portuguese data for 2005 and 2019 reveals that this inequality decrease is unambiguous and took place along the entire wage distribution. The application of this criterion can be seen in Figure A11.

B.3 Wage inequality decomposition: Between and Within

Wages are decomposed by construction as:

$$w_t^{i,j,f} = \overline{w_t} + (\overline{w_t^f} - \overline{w_t}) + (w_t^{i,j,f} - \overline{w_t^f}) \tag{A5}$$

where $w_t^{i,j,f}$ is the log of real hourly wages of worker i in firm f in year t, $\overline{w_t}$ is the average log of real hourly wage in the economy in year t, and $\overline{w_t^f}$ is the average log of real hourly wage in firm f (where worker i works) in year t. The wage of each worker can be seen as the sum of the average remuneration in the economy in that year, the difference paid on average by firms relative to the average wage in the economy, and the difference earned by workers relative to their firm's average wage. In order to obtain the within- and between-firms components of wage variance in each year, we rearrange and transform this identity into:

$$Var(w_t^{i,j,f} - \overline{w_t}) = Var(\overline{w_t^f} - \overline{w_t}) + Var(w_t^{i,j,f} - \overline{w_t^f}) \tag{A6}$$

Where $Cov(\overline{w_t^f} - \overline{w_t}; w_t^{i,j,f} - \overline{w_t^f}) = 0$ by construction[21]. Since wage variance is decomposed yearly, becomes a constant in Equation (A6) which once simplified is given by[22][23]

$$Var(w_t^{i,j,f}) = Var(\overline{w_t^f}) + \sum_{f=1}^{N} \omega_f Var(w_t^{i,j,f}|i \in f) \Leftrightarrow \tag{A7}$$

$$\underbrace{Var(w_t^{i,j,f})}_{Overall\ Inequality} = \underbrace{Var(\overline{w_t^f})}_{Between\ Firm\ Inequality} + \underbrace{\overline{Var(w_t^{i,j,f}|i \in f)}}_{Within\ Firm\ Inequality} \tag{A8}$$

This equation yearly decomposes the overall variance of log real hourly wages in a between-firms component (given by the variance across firm average wages), and a within-firm component (given by the weighted average of within-firm wage variance, with weight ω_f being the share of employment in firm f).

[21]Notice how $Cov(\overline{w_t^f} - \overline{w_t}; w_t^{i,j,f} - \overline{w_t^f}) = E([\overline{w_t^f} - \overline{w_t} - E(\overline{w_t^f} - \overline{w_t})][w_t^{i,j,f} - \overline{w_t^f} - E(w_t^{i,j,f} - \overline{w_t^f})]) = 0$

[22]To see this notice how: $\sum_i \frac{1}{N_F}(w_i - \overline{w_f(i)}) = \frac{1}{N_F}[\sum_i w_i - \sum_f n_f.\overline{w_f(i)}] = \frac{1}{N_F}[\sum_i w_i - \sum_i w_i] = 0$. Thus, $Var(w_t^{i,j,f} - \overline{w_t^f}) = \frac{1}{N_F}\sum_i (w_i - \overline{w_f(i)})^2$

[23]Furthermore, notice how: $\sum_{f=1}^{N_f} \omega_f Var(w_t^{i,j,f}|i \in f) = \sum_f \frac{n_f}{N_f}[\sum_{i(i \in f)} \frac{(w_i - \overline{w_f(i)})^2}{n_f}] = \sum_f \frac{n_f}{N}\frac{1}{n_f}[(w_i - \overline{w_f(i)})^2] = \frac{1}{N}\sum_f[(w_i - \overline{w_f(i)})^2] = Var(w_t^{i,j,f} - \overline{w_t^f})$

B.4 Balance in Terms of Region and Industry

Table A5: Balance Check for Matching, Industries and Regions

Sectors	*Industrial*	*Manufacturing*	*Construction*	*Retail*	*Hospitality*	*Others*
Control (%)	0.200	0.174	0.0360	0.277	0.113	0.200
Treated (%)	0.191	0.144	0.0410	0.255	0.176	0.192
Regions	*North*	*Algarve*	*Center*	*Lisbon*	*Alentejo*	*Islands*
Control (%)	0.276	0.0270	0.190	0.437	0.0300	0.0400
Treated (%)	0.314	0.0430	0.192	0.374	0.0370	0.0400

Sources: *IUTICE*, 2016 and 2018.
Note: This table displays balance checks for the control and treatment groups in terms of regions and industries.

B.5 Estimation Bias

In Section 5 we introduced the two-step procedure conducted by Alvarez et al. (2018) and Song et al. (2019). Such procedure starts by estimating a restricted wage model, in the vein of Abowd et al. (1999) that can be written in matrix notation as

$$w_t^{i,j,f} = D\delta^{f,P} + F\omega^{i,P} + B\tau^{j,P} + T\phi_t + \epsilon_t^{i,j,f} \tag{A9}$$

This equation is in everything equal to the one presented in Section 5 only making explicit the indicator vectors for the fixed effects. In this model, $w_t^{i,j,f}$ is a NT by 1 column vector with the wage outcome for each worker. D is a NT by F indicator matrix for firms. $\delta^{f,P}$ is a F by 1 column vector of relevant firm fixed effects parameters. F is an NT by N indicator matrix for workers and $\omega^{i,P}$ is a N by 1 column vector of relevant worker fixed effects parameters. B is a NT by J indicator matrix for job-titles, and $\tau^{j,P}$ is a J by 1 column vector of relevant job-titles fixed effects parameters. T is an indicator matrix for years (NT by T) and ϕ_t is a T by 1 vector of year-fixed-effects. Finally, $\epsilon_t^{i,j,f}$ is an NT by 1 column-vector of residuals. The model is said to be restricted for it omits time-varying observables. In the second step, the generated fixed effect values are used to determine the factors that influence on average the fixed effects and to decompose the variance of wages using the variance of these generated coefficients. However, the fact that these fixed effects are generated values deserves some precautions. Consider the firm fixed effect vector $\delta^{f,P}$. Notice that since $\hat{\delta}^{f,P}$ is a generated value, it is estimated with a certain degree of uncertainty:

$$\hat{\delta}^{f,P} = \delta^{f,P} + u_f^P \tag{A10}$$

Since $\hat{\delta}^{f,P}$ is used as a dependent variable in the second step of our procedure, it is critical to ensure that u_f^P is constant both cross-sectionally and across sub-periods. To answer the question of big of a bias is being imposed upon our estimates, one can leverage on the Frisch-Waugh-Lovell theorem (henceforth, FWL) and on the omitted variable bias formula. Start by defining an NT by 1 matrix of form

$$\Gamma = \begin{bmatrix} F & B & T \end{bmatrix} \begin{bmatrix} \omega^{i,P} \\ \tau^{j,P} \\ \phi_t \end{bmatrix}$$

So that our restricted AKM model can be written as

$$w_t^{i,j,f} = D\delta^{f,P} + \Gamma + \epsilon_t^{i,j,f} \tag{A11}$$

The annihilator ("residual maker") NT by NT matrix associated to Γ is defined as

$$M_\Gamma = I - \Gamma(\Gamma'\Gamma)\Gamma' \tag{A12}$$

where I is the identity matrix. Using the FWL theorem we can rewrite the **unrestricted** AKM model as:

$$M_\Gamma w_t^{i,j,f} = M_\Gamma D\delta^{f,P} + M_\Gamma X\beta \tag{A13}$$

where X is the initially left out vector of time-varying observables. We can express the above equation in matrix form. The normal equations boil down to:

$$\begin{bmatrix} D'M_\Gamma D & D'M_\Gamma X \\ X'M_\Gamma D & X'M_\Gamma X \end{bmatrix} \begin{bmatrix} \hat{\delta}^{f,P} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} D'M_\Gamma w_t^{i,j,f} \\ X'M_\Gamma w_t^{i,j,f} \end{bmatrix} \Leftrightarrow \tag{A14}$$

$$\begin{bmatrix} \hat{\delta}^{f,P} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} D'M_\Gamma D & D'M_\Gamma X \\ X'M_\Gamma D & X'M_\Gamma X \end{bmatrix}^{-1} \begin{bmatrix} D'M_\Gamma w_t^{i,j,f} \\ X'M_\Gamma w_t^{i,j,f} \end{bmatrix} \tag{A15}$$

Using the partitioned inverse result, we can write:

$$\hat{\delta}^{f,P} = (D'M_\Gamma D)^{-1} D'M_\Gamma w_t^{i,j,f} - (D'M_\Gamma D)^{-1} D'M_\Gamma X\hat{\beta} \tag{A16}$$

Since we are not estimating $\hat{\beta}$, $\hat{\beta} = 0$, thus

$$\hat{\delta}^{f,P} = (D'M_\Gamma D)^{-1} D'M_\Gamma w_t^{i,j,f} \tag{A17}$$

Furthermore, notice that the true model is given by

$$M_\Gamma w_t^{i,j,f} = M_\Gamma D\delta^{f,P} + M_\Gamma X\beta \tag{A18}$$

Combining these two equations, we obtain:

$$\hat{\delta}^{f,P} = (D'M_\Gamma D)^{-1} D'[M_\Gamma D\delta^{f,P} + M_\Gamma X\beta] \tag{A19}$$

Therefore, we can identify the bias u_f^P in:

$$\hat{\delta}^{f,P} = \delta^{f,P} + \underbrace{(D'M_\Gamma D)^{-1} D'M_\Gamma X\beta}_{u_f^P} \tag{A20}$$

The conditions under which this bias is null, or, at least, constant cross-sectionally and through time, are addressed in meta-analysis. Notice that if u_f^P is shown to be constant across both dimensions, then we have $u_f^P = u$. This constant will be absorbed by the constant in the second step, and the bias will therefore be innocuous for both the decomposition. Lewis and Linzer (2005) suggest using FGLS to solve this issue in the

	(1) 2005-2009	(2) 2010-2014	(3) 2015-2019
Mean Bias	1.51	1.56	1.58
SD Bias	0.12	0.11	0.13
Number of Firms	2000	1207	858

Table A6: Two-step procedure bias estimation

second-step. In this paper, instead, we use bootstrapped standard errors in the second step to fix this problem. Furtermore, in this paper, we estimate u_f^P using a random sample of firms and show that it is both constant across sub-periods and relatively constant cross-sectionally. The use of a random sample stems from the fact that since D is an NT observations by F firms matrix, inverting it is not computationally feasible given the size of our data. The results are shown in Table A6.

B.6 Data and Variables Details

Quadros de Pessoal Quadros de Pessoal is an administrative linked employer-employee job title dataset, for Portugal. The entity responsible for this statistical operation is the *Gabinete de Estratégia e Planeamento* (GEP) from the Ministry of Employment, Solidarity and Social Security (MTSSS), making the data available for Statistic Portugal (*Instituto Nacional de Estatística*). The panel 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 their employees and their earnings (for example gender of worker, highest education level completed, job titles, collective bargaining agreement, date of birth, occupation, date of hiring, and so forth), as well as information on the firms (for example, sector of activity, and so forth) and establishments. This requirement is meant as a way to verify if firms are complying with labor law. Since the employer is the one actually reporting the data, variables such as worker qualifications are less prone to measurement errors.

In terms of treating the data, each year, we first merge firm and worker data. Worker's observations having a worker ID with less than 6 digits or more than 10 digits are invalid and were therefore discarded. Whenever a worker appears twice within the same year in the panel with several jobs, his or her highest paying one was selected (since mostly likely, this is his or her primary job). Moreover, we keep, each year, observations for workers having: a job situation corresponding to dependent worker and at least 120 normal monthly hours of paid work (full-time workers). Each year, we also eliminate observations for workers without a complete basic remuneration, belonging to residual categories on job titles and 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. We also eliminate observations for 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 low coverage. Gross monthly earnings from dependent work are obtained by summing the earned remuneration of the worker and some irregular instalments too.

We use the consumer price index (CPI) deflator to convert nominal wages in real wages. After treating the datasets each year, the data is then appended and a panel is formed where each worker ID is tracked over time.

Regional Indicator Variable To identify the firms' (and workers') broad geographical regions, our setup relies on the Nomenclature of Territorial Units for Statistics at the regional level (NUTS 2). According to this classification, firms can be located in Lisbon, in the North, in Alentejo, in the Center region, in Algarve, in Madeira, or in the Azores. A broader nomenclature exists, NUTS 1, but its level of detail is coarser. Tables for NUTSII and NUTSI can be found on INE's website.

Education To determine the level of educational attainment of individuals, the paper focuses on a one-digit classification of highest educational attainment. The education labels were adjusted slightly for 2004 and 2005 to ensure a full harmonization of categories across time.

Sector Indicator Variable To determine the firms' sector of activity throughout the years, a crosswalk was used to adjust the classification in place before and after 2007. This was necessary since prior to 2007 activities were classified according to the *Classificação das Atividades Económicas Rev 2.1 (CAE Rev 2.1)*, but from 2007 onward, Portuguese activities have been revised to track international classifications and the new classification in place since then is the *Classificação das Atividades Económicas Rev 3 (CAE Rev 3)*. This harmonization crosswalk was built from the underlying two-digit CAE sectors and yielded 31 large categories, later reduced to 29 categories, once agriculture and fishing were discarded. For sake of reference, the coarser level of the classification used for economic sectors since 2007 is given by the sections on CAE Rev.3: A) agriculture, animal production, hunting, forestry and fishing (sector eliminated in our paper), 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 in our data because *Quadros de Pessoal* excludes employers of domestic service workers and people producing for own consumption).

Skill Composition Index To build our skill composition variable, we follow closely Lise and Postel-Vinay (2020). We start by creating a clean crosswalk between ISCO 2008 classification and SOC (Standard Occupational Classification). We then clean O*NET data so as to have a crosswalk between each one of the 35 skill dimensions and SOC codes. Next, we reduce the dimension of this matrix and make it a single vector. That is, we compute the first principal component using Principal Component Analysis (PCA). Call the principal component of each observation p_i. Equipped with this object, we normalize the principal component such that it is bounded between zero and one. Formally, let us denote S the set including each non-normalized principal component. We normalize each principal component according to

$$n_i = max\left\{\frac{p_i - min\{S\}}{max\{S\} - min\{S\}}; 0\right\}$$

Still using O*NET data, we convert 8 digit SOC codes into 6 digit SOC codes and adjust our skill measure so as to be the average of each 8 digit measure within each 6 digit code. For example, if profession 11111112 had a skill measure of 0.70 and profession 11111120 has a skill measure of 0.76, then profession 111111 will have a skill measure of 0.73. This leaves us with 747 different occupations. Table A7 shows the first ten and last ten entries of this crosswalk, as a sanity check for whether highly skilled professions indeed have a high skill measure associated to them. Once we have this, we merge this information of skills at the SOC occupation level with corresponding ISCO08 codes. We then, trim ISCO08 classification at the 3 digit level and take the mean of the skill measure within each of these 3 digits ISCO08 categories. This step is thus simply generating a correspondance between ISCO08 at the 3 digit level and an associated skill measure for each 3 digit occupational group. It is then possible to bring together ISCO08 data and the portuguese classification of Professions. Once we merge this information with Quadros de Pessoal, we are endowed with a measure of skill intensity for each worker in labor data. Averaging this measure within the firm, we get a measure of firm skill composition.

Table A7: Skill Measure by Occupational Group

Highest Skill levels		
111011	1,000	Chief Executives
192012	0,869	Physicists
119151	0,844	Social and Community Service Managers
119121	0,838	Natural Sciences Managers
212011	0,826	Clergy
193032	0,824	Industrial-Organizational Psychologists
291067	0,810	Surgeons
113131	0,807	Training and Development Managers
113121	0,805	Human Resources Managers
172051	0,798	Civil Engineers
Lowest Skill levels		
513023	0,079	Slaughterers and Meat Packers
372012	0,076	Maids and Housekeeping Cleaners
537111	0,074	Mine Shuttle Car Operators
372011	0,071	Janitors and Cleaners, Except Maids and Housekeeping Cleaners
473015	0,071	Helpers–Pipelayers, Plumbers, Pipefitters, and Steamfitters
359021	0,061	Dishwashers
516021	0,020	Pressers, Textile, Garment, and Related Materials
452041	0,016	Graders and Sorters, Agricultural Products
537061	0,007	Cleaners of Vehicles and Equipment
537064	0,000	Packers and Packagers, Hand

Sources: O*NET Dataset, ISCO classification and National Classification of Portuguese Occupations.
Note: This table reports the skill composition measure built in this paper associated with selected occupations. We select the ten highest ranked and the ten lowest ranked occupations and display the associated skill score index.