Education, Sorting and Wages: A Structural Matching Approach

Pauline Corblet*

August 19, 2025

Abstract

This paper proposes a novel empirical framework for capturing drivers in the wage structure. It generalizes the one-to-one random utility matching framework to a one-to-many setting, and enables the estimation of workers' non-wage amenities and firms' productivity using observed matching and wage distributions. Crucially, the one-to-many framework allows identification of cross-complementarity terms between workers in both non-wage amenities and productivity, in addition to the standard sorting between workers and firms. This makes it possible to identify how worker segregation impacts wage gaps, along with worker and firm sorting. The paper applies the framework to the Portuguese labor market, and shows that the compression of the wage structure along the education dimension is not only driven by an education expansion, but also by a reduction in complementarities in productivity between educated workers.

Keywords: One-to-Many Matching, Sorting, Educational Changes
JEL Codes: C78, J24, J33

*New York University Abu Dhabi, pauline.corblet@nyu.edu
This work is based on my Ph.D job market paper. I thank my advisor Alfred Galichon for his valuable guidance and support. My Ph.D was partly financed through the ERC grant EQUIPRICE. I acknowledge the IAAE conference travel grant. I am indebted to Francesco Franco who generously helped me navigate the application process to obtain the data used in this project. I also thank Jean-Marc Robin, Moshe Buchinsky, Florian Oswald, Kerstin Holzheu, Zsofia Barany, Arnaud Dupuy and Etienne Wasmer as well as all the participants to various conferences and seminars for their comments and feedback. All remaining errors are mine.

1 Introduction

Over the past four decades wage structures have shifted in markedly different directions across advanced economies. In the United States, pay gaps between college-educated and less educated workers have widened (Goldin and Katz (2008)), while several European countries have seen those gaps stagnate (as shown by Krueger et al. (2010) for Germany and Italy) or even compress (as shown by Verdugo (2014) for France). Gender differentials have also evolved unevenly, closing rapidly in some countries, and remaining large in others (Krueger et al. (2010)). These diverging patterns remind us that who earns what relative to whom is the result of a complex interplay between supply of skills, firms' productivity structure, technological change, and worker preferences. While standard explanations focus on linear technological change and relative worker supply to explain changes in the wage structure, this paper proposes an approach that incorporates non-wage amenities and interdependence between workers employed in the same firm. The comprehensive empirical framework developed here rests on a workers-to-firm matching model that produces rich sorting patterns between workers and firms, as well as within the workforce, and accounts for both productivity and non-wage amenities. Standard explanations for wage differentials focus on the role of educational attainment, which changes the relative supply of educated workers[1]; technological change, which describes the relative increase in educated workers' productivity; and worker-firm sorting, whereby workers are employed in firms with different productivity levels based on their education or gender. In addition to these three channels, the framework developed here incorporates non-wage amenities, and interdependence patterns between workers employed in the same firm. In line with the theory of compensating differentials, workers who enjoy high non-wage amenities in a given firm are willing to accept lower wages to work in that firm. Therefore, if educated workers get higher non-wage amenities than less-educated workers, it puts a downward pressure on the education wage premium that is proportional to the amenities gap. Interdependence between workers within the same workforce means the level of non-wage amenities depends not only on the employer, but also on the other employees. It also translates into complementarity or substitutability between workers in productivity.

This paper builds a static one-to-many matching model with transferable utility, in which a single firm matches with several workers, referred to together as its workforce. The model can be summed up as a generalization of Choo and Siow (2006) one-to-one framework. Workers differ along two observable dimensions - education level and gender - which are

[1] Educated refers to either secondary (high school) or tertiary education graduates. Less-educated refers to individuals who did not complete high school.

summed up in worker types. The workforce composition counts the number of workers of each type employed. To capture differences between firms, I use Bonhomme et al. (2019)'s method, which summarizes firm heterogeneity into one-dimensional classes. Applying the method results in firm types that capture firm heterogeneity based on their wage distribution. Firm types are then ranked by average wage from the lowest to the highest type. When a firm matches with a workforce, they create a match surplus that depends on the firm's type and the workforce composition in terms of worker types. This surplus can be separated into a productivity component, that accrues to the firm, and an amenity component that accrues to each member of the workforce. The amenities capture workers' non-wage preferences towards each firm type. The utility is transferable in the form of wages that the firm pays to the workers in its workforce. Firms seek to maximize total profit, which is additive in the difference between productivity and total wage bill, plus a random shock. Workers maximize their utility, which is additive in amenities, wages, and a random shock. At equilibrium, wages clear the market, and each agent matches with their best option given wages. The model can generate a rich distribution of wages that depends both on workers' observable characteristics and the firm's type, as well as on the employed workforce. Two workers of the same education level, age and gender employed by the same firm type can be paid different wages if the firms they work in employ different workforces. The model also predicts the equilibrium matching, that is, the joint distribution of firms and workforces. Prediction of both wage and matching distributions makes it possible to identifying productivity and amenities separately. Together they increase the chances of matching, but amenities lower wages while productivity increase them. The model allows for a flexible parametrization of productivity and non-wage amenities, that depends not only on individual firm and worker types, but also on workforce composition. A worker's contribution to productivity depends positively or negatively on the other worker types hired by the firm. Workers' non-wage amenities also depend on the other workers in the workforce.

The first paper to explore one-to-many matching in the labor literature is Kelso and Crawford (1982). It proposes an algorithm that yields a stable matching and corresponding wages, under the condition that workers are gross substitutes. This condition restricts the firm's production function by imposing that workers substitute each other. In the matching literature, a similar condition is traditionally imposed on buyers in many-to-one assignment problems (Bikhchandani and Ostroy (2002), Vohra (2011)), but significant progress has been made in relaxing this condition. Azevedo and Hatfield (2018) show that under transferable utility, substitutable preferences are not necessary to guarantee the existence of a stable matching in large markets. The present paper's proof of existence of an equilibrium matching rests on a similar argument[2]. In the labor literature, recent work has studied the matching of multi-dimensionally skilled workers to jobs. Examples include Lindenlaub (2017) who study the implications of technological change on multidimensional matching, and Lise and Postel-Vinay (2020) who incorporates multidimensional matching into a search model[3]. The model built in the present paper abstracts from search by capturing frictions by a random utility term. It also features multidimensional matching, and accounts for workers' interaction within firms, which those papers cannot do by matching workers to jobs. Another strand of the literature incorporates workers' non-wage amenities in models of matching, as does the present paper. Lindenlaub and Postel-Vinay (2021) studies the observable characteristics that drive these non-wage amenities. They use worker's mobility between firms to identify them, in the same vein as Sorkin (2018) who estimates that amenities account for more than half of the firms' contribution to earnings variance. Lamadon et al. (2024) also decompose wage dispersion between non-wage amenities and search frictions. The approach to identification is slightly different in the present paper, as it uses observed matches instead of moves. Finally, Choné and Kramarz (2021) uses a many-to-many matching model to study the effects of skill 'bundling', that is, that a single worker displays different levels of skills in different tasks.

One of the chief advantages of random utility models *a la* Choo and Siow (2006) is that they can easily be brought to the data, and the present model is no exception. Under a standard Extreme Value 1 assumption on the distribution of the firms' and workers' random shocks, both non-wage amenities and productivity are non-parametrically identified using observed matching and wage distribution. Intuitively, total surplus, which is the sum of the workforce productivity at a given firm type and all workers' non-wage amenities at that firm type is identified by the strength of matching. A high total surplus yields many matches in the data. Next, productivity and non-wage amenities have opposite effect on wage: the former correlates positively with wage, but the latter correlates negatively. This enables us to tell productivity and non-wage amenity apart. The firm productivity and worker non-amenities can be flexibly parametrized, as functions of the firm type and workforce composition. The parametrization is additive in own-productivity terms, that depend linearly and log-linearly on the number of workers of a given education level and gender employed by the firm, and cross-productivity terms, that capture interaction terms between different worker types within the workforce. The model estimation requires observation of moments from the matching and wage distribution and is performed in two steps. The first step estimates the

[2] Che et al. (2019) established a similar result in a framework with non-transferable utility. Their paper contains an enlightening example that provides intuition on the role of large markets.

[3] See also Lise et al. (2016) for a classic model of search and matching with uni-dimensional skill.

total match surplus (the sum of productivity and non-wage amenities) parameters by fitting matching moments. In the context of the standard one-to-one matching model, Galichon and Salanié (2024) show that these moments can be fitted through a Poisson regression, and the present paper extends this result to the one-to-many case. The second step estimates the non-wage amenities parameters by fitting wage moments with a generalized method of moments (GMM) procedure.

Since the 1970s, many economies have undergone large education expansions. As a result, the ratio of educated workers (whether secondary or college-educated) to less educated workers present in labor markets has increased. This occurred in the context of a race between education and technology (Goldin and Katz (2008)): the supply of educated workers was growing, but demand for their skill was growing too. The consequences of this race on the wage structure differ by country. In the United States, college-educated workers have enjoyed a rising wage premium throughout the 1980s and 1990s, but this trend has now plateaued. Contrary to the United States, some European countries, such as Portugal or France, faced a wage compression along the education dimension. This paper focuses on Portugal, which is a striking case of education expansion: between the early 1990s and late 2020s, the share of high school graduates quadrupled, and the share of university graduates[4] went from less than 5% to one fifth of the working population. Over the same period as this sizable education expansion, the university and high school wage premia -referred together to as 'education wage premia'- declined substantially. A classical explanation is that technology 'lost' the race to education (Goldin and Katz (2008)): the increase in relative supply could have overtaken the increase in relative demand. I check for this hypothesis by simulating the matching and wage distribution in the absence of the education expansion in Portugal, and find that the education expansion is indeed a key driver of the fall in the education wage premia. The model proposed in this paper also allows for a complex structure for firms' productivity functions, which account for firm heterogeneity and changing interdependency patterns between workers. Counterfactuals show that non-linear changes in productivity and uneven skill-biased technological change are also driving the decline of the education wage premium. Many emerging economies are still experiencing a rapid growth in the ranks of young educated workers. Therefore, it is crucial to understand how these dynamics unfolded in economies to help forecast future trend and inform the development of effective public policies.

Beyond this crucial finding, the model's estimation yields the following main results. First,

[4]University education refers to any tertiary degree, which may also include short vocational degrees completed after high school. Individuals who graduate from a tertiary education institution are referred to as university graduates throughout the paper.

the higher the firm type, the more productive the workers, which confirms the literature's interpretation of firm types as firm heterogeneity in productivity. However, if high school and university graduates are more productive less-educated workers in high-type firms, they are actually less productive in low-type firms. This observation suggests the nature of tasks differ in low-type and high-type firms, and human capital acquired through education does not fit well low-type tasks, either because of technology complementarities or over-education. Over time, educated workers' productivity increased in low- and medium-type firms. Therefore, the first result from the model is that skill-biased technological change did occur in Portugal between the 1990s and 2010s, but unevenly across firms. Second, all workers' non-wage amenities are negative, implying a disutility from working. The higher the firm type, the larger the disutility. This result is consistent with higher job demands in high-type firms. A larger disutility implies an upward pressure on wages, hence the higher wages offered in high-type firms do not stem only from higher productivity but also compensating differentials. In the early 1990s less-educated workers' disutility is larger than that of educated workers', but the opposite is true in the late 2010s. The rich model parametrization allows marginal returns to an additional worker of a given type to vary by firm type and workforce composition. Two worker types are said to be complements if the marginal returns in productivity of one type are increasing in the other, and substitute if they are diminishing. Model estimation shows approximately one-third of worker pairs are always substitutes, one third are always complements, and the last third varies depending on workforce composition. This distribution is approximately constant over time. A closer look at these interdependence patterns shows they are strongest among university graduates pairs in the early 1990s: male university graduates strongly dislike working with other male university graduates, but are more productive if they are well represented in the workforce. I then use counterfactuals to identify the forces driving the education wage premia's decline. There are two main results to highlight. First, both education wage premia, and especially the high school wage premium, would have been much higher in a world where education share had not changed between 1991 and 2017, showing that the education expansion is a key driver of the education wage premia decline. Second, technological change in the productivity parameters had a ambiguous effect on the education wage premia. On the one hand, skill-biased technological change in the linear productivity parameters drives both education wage premia up. On the other hand, changes in cross-productivity parameters actually exert a downward pressure on the university wage premium. Indeed, In the early 1990s, educated workers are strong complements to each other and other education levels in high-type firms, i.e. the marginal returns in firm productivity of educated workers are positive if they are already well-represented in the workforce. This strong complementarity fades out over time, which brings university graduates' wages down. To understand how complementarities between workers relate to the gender wage gap, I run a last counterfactual in which I fix women's interdependence patterns in productivity to be the same as men's over time. I find that this counterfactual leads to much smaller gender gap in 1991, and an inverted gender wage gap in 2017: women earn more than men. This shows the gender pay gap is driven by strong complementarity patterns within educated men, and not solely compensating differentials in non-wage amenities, as is often pointed out in the literature.

Another strand of the literature focuses on the role of firms versus workers fixed effects in explaining wage dispersion, starting with Abowd et al. (1999). Recently, this literature has focused on the role of worker-firm sorting and worker segregation to explain wage dispersion. Card et al. (2018) use the same Portuguese matched employer-employee dataset used here and showed that 40% of the observed difference in wage across firms is due to worker-firm sorting. Song et al. (2019) decompose between-firm wage dispersion in the US into sorting and segregation. Lamadon et al. (2022) also decompose wage dispersion using a model with imperfect competition. Babet et al. (2022) also underscore the importance of worker segregation in France. The present paper contributes to this literature by providing a structural framework with which to think about worker and firm sorting, and worker segregation within workforce. The paper takes a slightly different approach by estimating discrete firm fixed effects (Bonhomme et al. (2019)), but instead of focusing on worker fixed effects it studies how worker interact within the firm based on observable characteristics, namely education and gender.

Finally, the present paper contributes to the literature on education wage premia, which have been studied extensively in recent decades. Katz and Murphy (1992) and Card and Lemieux (2001) show that the rising US college wage premium in the 1980s and 1990s is driven by a combination of skill-biased technological change, combined with a slow-down in educational attainment. Starting in the 2000s, the college wage premium is stagnating in the US. Different explanations for this fact have been put forward, either related to declining demand for skill (Beaudry et al. (2015)) or the growing supply of college graduates (Blair and Deming (2020)). In Portugal, Centeno and Novo (2014) argued that wage inequality has increased over the last thirty years, but do not evidence the decreasing education wage premia. Cardoso (2004) studies the fit between workers' education and their jobs between 1986 and 1999, and finds that it improved during that period. The present paper offers an update on these observations, which have reversed since.

Section 2 describes the one-to-many matching model used to model the labor market. Section 3 presents the model's estimation strategy, Section 4 discusses the estimation results, Section 5 presents the model counterfactuals, and Section 6 concludes.

2 The Model

Employer employees match formation is modeled using a two-sided random utility framework, in the same spirit as Choo and Siow (2006). The key generalization of this paper is that the one-to-one matching model is extended to a one-to-many matching model, whereby a single firm can employ multiple workers. This innovation allows me to capture rich interdependence patterns between workers who are employed by the same firm, both in the firm's production function and in the workers' non-wage utility. Key features of the framework are heterogeneous workers and firms, transferable utility between firms and workers, surplus separability into a systematic and a random part, and the absence of search frictions.

2.1 Set Up

There is a large number of workers and firms who match up. An individual worker is denoted by i and an individual firm by j. Workers are characterized by their type, which can stem from any trait observable to the analyst: education gender, experience, etc. A worker type is denoted by $x \in \mathcal{X}$, where $\mathcal{X}$ is finite, and its cardinal is X. The total mass of workers of Type x present in the labor market is m_x. Similarly, firms are characterized by their type $y \in \mathcal{Y}$, where $\mathcal{Y}$ is finite with cardinal Y. The mass of Type y firms is n_y. A single firm can employ (or match with) multiple workers. These workers form the firm's workforce. The workforce composition, or type k counts the (finite) number of workers of each Type x employed in the workforce. That is, k is a vector of size X:

$$k = (k_1, \dots, k_X) \tag{1}$$

and each k_x is equal to the number of type x workers employed in the workforce. For all x, k_x is an integer between 0 (if no worker of type x is employed) and n_x (if all type x workers are employed). Every workforce k lives in $\mathcal{K}$, the cartesian product of the sets $\{0, \dots, n_x\}$, excluding the empty workforce:

$$k \in \mathcal{K} = (\Pi_{x\in\mathcal{X}}\{0, \dots, n_x\}) \setminus \{0, \dots, 0\}^{5}.$$

The set $\mathcal{K}$ is finite. In practice it is represented by a grid of workforces k.

[5] The matching could also occur on a subset of allowed workforces $\mathcal{K}^c \subset \mathcal{K}$, and all the arguments made below would go through.

Worker utility and firm output are additive in a systematic and a random part. Worker i of type x obtains systematic utility u_{xyk} when matched to a firm of type y within a workforce k, where

$$u_{xyk} = A_{xyk} + w_{xyk}. \tag{2}$$

A_{xyk} is a non-wage amenity: it accrues to the worker just from being employed in firm y and workforce k. There is no sign restriction on A_{xyk}, so that it could also represent a disutility if negative. The term w_{xyk} is the monetary transfer from the firm y to the worker x. It depends on the type of workforce k employed by the firm. Workers also have the option to remain unemployed, in which case their systematic utility is normalized to 0. The firm's systematic profit is

$$v_{yk} = \Gamma_{yk} - \sum_{x \in \mathcal{X}} k_x w_{xyk}. \tag{3}$$

Γ_{yk} is the type y firm's total output when employing workforce k. The firm transfers w_{xyk} to each of its workers, which sum up to a $\sum_{x \in \mathcal{X}} k_x w_{xyk}$ total wage bill. The total systematic surplus from a match between a firm and a workforce is the sum of the workers' utility and the firm's profit:

$$\Phi_{yk} = \sum_{x \in \mathcal{X}} k_x u_{xyk} + v_{yk} = \sum_{x \in \mathcal{X}} k_x A_{xyk} + \Gamma_{yk} \tag{4}$$

where transfers cancel out: utility is perfectly transferable between firm and workers in the workforce.

In addition to their systematic utilities, workers and firms also experience a random taste shock, which captures the unobservables that are not reflected by types x and y but still contribute to the match formation. For instance, workers' gender and education level are usually observed by researchers, so they would enter the workers' observable type x. However, an individual worker's taste for working in a specific industry, and/or a small or large workforce are unobserved, so these are accounted for by the random shocks. Therefore, worker i's total utility is

$$u_{x_i yk} + \varepsilon_{iyk} \tag{5}$$

or ε_{i0} if they are unemployed, and firm j's total profit is

$$v_{y_j k} + \eta_{jk}. \tag{6}$$

Both the worker's and the firm's shocks are specific to the individual worker or firm drawing them. The worker's shock depends only on the employer's type and workforce composition (not on the identity of the firm or the co-workers), and the firm's shock depends only on the workforce composition. The shocks are drawn before matching occurs and are common knowledge. The random shocks satisfy the following assumption.

Assumption 1. *Random shocks satisfy the following:*

(i) For each pair of two workers i and i', ε_{iyk} and $\varepsilon_{i'yk}$ are mutually independent and identically distributed.

(ii) For each pair of two firms j and j', η_{jk} and $\eta_{j'k}$ are mutually independent and identically distributed.

(iii) For a worker i and a firm j, ε_{iyk} and η_{jk} are mutually independent.

(iv) ε_{iyk} is independent of α_{x_iyk}, η_{jk} is independent of γ_{yk}.

(v) $(\varepsilon_{iyk})_{y,k}$ and $(\eta_{jk})_k$ are distributed as extreme value 1 (Gumbel distribution)

The Gumbel assumption is in line with the literature that deals with heterogeneity in matching models (Choo and Siow (2006), Galichon and Salanié (2024)), although as shown in Galichon and Salanié (2022) in the one-to-one case, identification is possible with a general class of distributions.

A matching market is characterized by exogenous distributions of worker and firm types $m = (m_x)_{x\in\mathcal{X}}$ and $n = (n_y)_y$, amenities and production $\alpha = (\alpha_{xyk})_{x\in\mathcal{X},y\in Y,k\in\mathcal{K}}$ and $\gamma = (\gamma_{yk})_{y\in\mathcal{Y},k\in\mathcal{K}}$, and the distribution of stochastic shocks ε and η. The next subsection describes the workers' and firms' maximization problem and the resulting competitive equilibrium.

2.2 Competitive Equilibrium

Type x workers supply a mass S^x_{yk} to Type y firms and workforce k, while Type y firms demand a mass D^y_k of workforces k. Let S^x_0 be the supply of unemployed workers. Note that S^x_{yk} and D^y_k are not in the same units: the former counts workers, while the latter counts workforces. To count the mass of Type x workers demanded by firms, it is necessary to multiply D^y_k by k_x. Both supply and demand are determined by the workers and firms' best option on the matching market.

Definition 1. *The supply S^x_{yk} of Type x workers willing to be employed in Type y firms and workforces k is*

$$S^x_{\bar{y}\bar{k}} = m_x \mathbb{P}\left[(\bar{y}, \bar{k}) = \underset{y\in\mathcal{Y},k\in\mathcal{K}}{\arg\max} \{u_{xyk} + \varepsilon_{yk}, \varepsilon_0\}\right]$$

The supply of Type x workers willing to remain unmatched is

$$S_0^x = m_x \mathbb{P}\left[0 = \arg\max_{y\in\mathcal{Y},k\in\mathcal{K}} \{u_{xyk} + \varepsilon_{yk}, \varepsilon_0\}\right]$$

The demand D_k^y from Type y firms for workforce k is

$$D_{\bar{k}}^y = n_y \mathbb{P}\left[\bar{k} = \arg\max_{k\in\mathcal{K}} \{v_{yk} + \eta_k\}\right]$$

Let $S = \left(\left(S_{yk}^x\right)_{x\in\mathcal{X},y\in\mathcal{Y},k\in\mathcal{K}}, \left(S_0^x\right)_{x\in\mathcal{X}}\right)$ and $D = \left(D_k^y\right)_{y\in\mathcal{Y},k\in\mathcal{K}}$.

Definition 1 defines supply and demand through workers' utility and firms' profit maximization problems. A worker is part of the supply S_{yk}^x if they are of Type x and their best option is to be employed by a firm y within workforce k. Note that a worker chooses both a firm type and a workforce. A firm is part of demand D_k^y if it is of Type y and its best option is to employ workforce k. The supply of unemployed workers S_0^x is the mass of Type x workers whose best option is to remain unmatched.

Under Assumption 1, S and D write in closed-form as functions of u and v.

Proposition 1. *Under Assumption 1, the equilibrium supply and demand satisfy*

$$\begin{aligned}
S_{yk}^x &= n_x \frac{\exp(u_{xyk})}{1 + \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \exp(u_{xyk})} \\
S_0^x &= n_x \frac{1}{1 + \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \exp(u_{xyk})} \\
D_k^y &= m_y \frac{\exp(v_{yk})}{\sum_{k\in\mathcal{K}} \exp(v_{yk})}
\end{aligned}$$

Proof. The proof consists in spelling out the choices probabilities of each worker and firm type. The specific functional forms stem from Assumption 1. The full proof is in Appendix A. □

Proposition 1 shows supply and demand take familiar logit forms under assumption 1, similar to the probabilities described in Choo and Siow (2006). Both depend on the wages underlying utilities u and profits v. Using supply and demand, we can now define formally the competitive equilibrium.

Definition 2. *A competitive equilibrium in this model is a set of underlying wages $w = (w_{xyk})_{x\in\mathcal{X},y\in\mathcal{Y},k\in\mathcal{K}}$, supply S and demand D such that excess demand is zero; that is*

$$D_k^y - k_x S_{yk}^x = 0 \quad \forall x \in \mathcal{X}, y \in Y, k \in \mathcal{K}\,.$$

In other words, Definition 2 states that at equilibrium, the mass of workforces k demanded by firms y is equal to what each worker type x is ready to supply to y and k. Note that excess demand has to be zero for all worker types x: each worker type x has to supply their required number k_x in workforce k to the demand from y. Meanwhile the demand from y applies only to workforces k, which is a given mix of worker types x, and not toindividual worker types x. At equilibrium, supply and demand are referred to as matching.

Definition 3. *The matching distribution* $\mu = (\mu_{yk})_{y\in Y, k\in\mathcal{K}}$ *is equal to the equilibrium demand* D*:* $\mu_{yk} = D_{yk}$ *for all* $y \in \mathcal{Y}$ *and* $k \in \mathcal{K}$*. The unemployed distribution* $\mu_0 = (\mu_{x0})_{x\in\mathcal{X}}$ *is equal to the equilibrium supply* S_0^x*.*

Proposition 1 and Definition 2 do not offer a practical way of computing the competitive equilibrium, which is needed to bring the model to the data. The next subsection describes how equilibrium matching and wages can be computed from a social planner problem.

2.3 The Social Planner Problem

This section shows that the matching in the competitive equilibrium defined in Definition 2 is the solution to a social planner problem. The social planner problem is a strictly concave maximization problem under constraints that are the margin equations of the matching distribution. The equilibrium matching is therefore guaranteed to exist and is unique. Under Assumption 1 the social planner problem provides closed-form expressions for equilibrium matching and wage. We start with the definition of workers' and firms' expected indirect payoffs that appear in these closed forms.

Definition 4. *Type x worker's and Type y firms expected indirect utilities/profit are*

$$G(u_{x\cdot}) = \mathbb{E}\left[\max_{y\in\mathcal{Y}, k\in\mathcal{K}} \{u_{xyk} + \varepsilon_{yk}, \varepsilon_0\}\right]$$
$$H(v_{y\cdot}) = \mathbb{E}\left[\max_{k\in\mathcal{K}} \{v_{yk} + \eta_k\}\right]$$

Under Assumption 1, G_x and H_y write in closed-form as a function of $u_{x\cdot}$ and $v_{y\cdot}$.

Proposition 2. *Under Assumption 1, expected utilities and profits from Definition 4 are*

$$G_x(u_{x\cdot}) = \log\left(1 + \sum_{y\in\mathcal{Y}, k\in\mathcal{K}} \exp(u_{xyk})\right)$$
$$H_y(v_{y\cdot}) = \log \sum_{k\in\mathcal{K}} \exp(v_{yk})$$

Proof. In Appendix A ☐

Building on Definition 4 and Proposition 2, Theorem 1 below states a first welfare equivalence: the competitive equilibrium matching is the solution to social planner problem. The problem's objective function is additive in two parts: the expected total surplus over the matching distribution, and a penalty that reflects the frictions introduced by the random part of workers' utility and firms' profit. That penalty is the expected value of the shocks over the population, which we can compute in closed-form under Assumption 1. The margin equations are the constraints of the social planner problem, which ensure that the total number of workers of each type x employed in the workforce is equal to the mass of workers of that type n_x, and that the total number of workforces k employed by each firm type y is equal to the mass of firms of that type m_y.

Theorem 1. *The equilibrium matching μ is the solution to the social planner problem*

$$\max_{\mu} \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \mu_{yk}\Phi_{yk} - \mathcal{E}\left(\mu, n, m\right) \tag{7}$$

such that the margin equations

$$\sum_{y\in\mathcal{Y},k\in\mathcal{K}} k_x\mu_{yk} = n_x \quad \forall x \in \mathcal{X}, \quad \sum_{k\in\mathcal{K}} \mu_{yk} = m_y \quad \forall y \in \mathcal{Y}, \tag{8}$$

are satisfied. Under Assumption 1,

$$\mathcal{E}\left(\mu, n, m\right) = -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y},k\in\mathcal{K}} k_x\mu_{yk} \log\frac{k_x\mu_{yk}}{n_x} - \sum_{x\in\mathcal{X}} \mu_{x0}\log\frac{\mu_{x0}}{n_x} - \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \mu_{yk}\log\frac{\mu_{yk}}{m_y}$$

and the equilibrium matching exists and is unique.

Proof. The competitive equilibrium conditions derived in Proposition 1 are the first-order conditions to the social planner problem (7). For a complete derivation of the social planner problem, see Appendix A. ☐

Proposition 3. *Under Assumption 1, the equilibrium matching and wage distribution have closed-form expressions.*

- *The matching* $\mu = (\mu_{yk})_{y,k}$ *writes*

$$\log \mu_{yk} = \frac{1}{1+\sum_x k_x}\left(\Phi_{yk} - \sum_x k_x \left(G_x(u_{x.}) - \log n_x\right) - \left(V_y(v_{y.}) - \log m_y\right) \right. \\ \left. - \sum_x k_x \log k_x \right) \tag{9}$$

$$\text{and } \log \mu_{x0} = G_x(u_{x.}) - \log n_x$$

- *The wage* $w = (w_{xyk})_{x,y,k}$ *writes*

$$w_{xyk} = G_x(u_{x.}) - \log n_x - A_{xyk} + \log k_x + \log \mu_{yk} \tag{10}$$

Proof. Start from the equilibrium equations in Proposition 1 and take logs on each side to obtain

$$\log \mu_{yk} = \log n_x - \log k_x + u_{xyk} - G_x(u_{x.}) \quad \forall x, y, k$$
$$\log \mu_{x0} = \log n_x - G_x(u_{x.}) \quad \forall x$$
$$\log \mu_{yk} = \log m_y + v_{yk} - H_y(v_{y.}) \quad \forall y, k$$

where $u_{xyk} = A_{xyk} + w_{xyk}$ and $v_{yk} = \Gamma_{yk} - \sum_{x \in \mathcal{X}} k_x w_{xyk}$. Multiply the first line by k_x, sum over all x and add the third line to obtain (9). Equation (10) obtains from the first line. □

In practice, one defines worker type fixed effects $U_x = G_x(u_{x.}) - \log n_x$ for all $x \in \mathcal{X}$ and firm type fixed effects $V_y = H_y(v_{y.}) - \log m_y$ for all $y \in \mathcal{Y}$. The equilibrium matching μ is obtained for the fixed effects $U = (U_x)_x$, $V = (V_y)_y$ that solve the margin equations (8).

The closed-form expressions in Proposition 3 provide clear insights into the mechanisms of the model. In Equation (9), the larger the total match surplus Φ, the more frequent the occurrence of the match. This is relative to the sum of expected indirect payoffs of all agents in the match $\sum_x k_x G_x(u_{x.}) + V_y(v_{y.})$. $G_x(u_{x.})$ and $V_y(v_{y.})$ are both equilibrium objects as u_{xy} and v_{xy} depend on w. The match occurrence also depends positively on the masses of workers and firms n_x and m_y, which act as scaling factors. Finally, the size of the workforce acts as a penalty on the match occurrence through $\sum_x k_x \log k_x$. From equation (10), a higher total match surplus also yields a higher wage, through the matching term $log \mu_{yk}$. On the other hand, high non-wage amenities A_{xyk} compared to expected indirect payoffs $G_x(u_{x.})$ reduce the wage: a worker who is happy to work for a given firm type in a given workforce accepts a lower wage all other things being equal. A worker's wage also depends positively on $\log k_x$, the logarithm of the number of workers of type x employed in the workforce, and negatively on the total mass of workers n_x, reflecting that scarcity allows workers to obtain higher wages.

Proposition 3 also offers clear insights on the identification of amenities and production functions: observation of the matching distribution μ identifies total match surplus Φ, up to worker and firm type fixed effects. Assuming amenities in unemployment are zero, observing the unemployed μ_0 identifies the worker type fixed effects U. However, only observation of vacant firms, which are not modeled here[6], could identify the firm type fixed effects V. As a result, total match surplus is only identified up to a constant in firm type. Given that total surplus is identified using matching and unemployed distribution, and fixed effects are set by the margin equations, the wage distribution identifies the amenities A_{xyk} (Eq. (10)). Output Γ_{yk} is then simply the difference between total surplus Φ_{yk} and the sum of amenities A_{xyk}. Two remarks are in order. First, the identification discussed here is non-parametric, in the sense that Φ, A and Γ have not been given functional forms. Second, the non-parametric identification constrains the surplus parametrization, in that we know we cannot identify firm fixed effects. On should also keep in mind the interpretation of the amenities and production is always relative: for workers it is relative to the outside option of being unemployed, and for firms to the other workforces they could employ.

3 Estimation

3.1 Data

The model is estimated on Portuguese matched employer-employee data *Quadros de Pessoal* yearly from 1991 to 2017. The data were obtained from a yearly mandatory survey that covers all firms, run by the Instituto Nacional de Estatística (INE). It covers all employees in the private sector (except domestic workers). These data are supplemented it with publicly available yearly unemployment rates[7]. The data provide information on workers' gender and education level, and allow for hourly earnings computation. Estimation is performed on full-time and unemployed workers only, between 16 and 64 years old. Table 1 summarizes the sample size. Over the 26 years of data, almost 8 million individual workers are observed, for an average of about 1.8 million per year. They are employed in more than 4 hundred thousand different firms over that time period, and about 1.3 hundred thousand firms are observed each year. For more details on the data and cleaning, see Appendix C.

[6]Modeling vacant firms poses no theoretical issues, they are not part of this framework because the data does not provide any information on them

[7]available from the World Bank data website

Table 1: Data Summary

	Total number	Average number per year
Workers	7,967,982	1,770,053
Firms	539,037	132,227

3.2 Parametrization

Estimating the matching model requires parametrizing worker and firm types. Worker types are set as a combination of their gender and their education level. Given the two genders and three education levels, there are six worker types:

$$x \in \{\text{F-NoHS}, \text{F-HSG}, \text{F-UG}, \text{M-NoHS}, \text{M-HSG}, \text{M-UG}\}$$

The first letter, M or F, refers to the worker's gender (male or female) and the acronyms NoHS, HSG and UG refer to the worker's education level (no high school for less-educated workers, high school Graduates, and university graduate).

Firm types are established using the group classification method from Bonhomme et al. (2019). This method is standard in the literature to capture firm heterogeneity in productivity. It allows firm heterogeneity to be summarized along a single dimension. The model can accommodate several dimensions, such as industry, location or maturity, but using a single dimension in estimation reduces the computational burden and eases interpretation. The method's core idea is to group firms into discrete heterogeneity bins based on their wage distribution using the *kmeans* algorithm. To accomodate the present's study time frame, the method is ran on firms in 2017 and use the resulting centroids to allocate firms to firms in previous years. Types are then ranked by average wage. The discrete heterogeneity groups are then used as firm types y in the model's estimation. The full procedure is described in Appendix C and Section 4 discusses the resulting firm types in more detail.

Given worker and firm types, non-wage amenities are parametrized as follows:

$$A_{xyk} = \alpha^1_{xy} + \sum_{x' \in \mathcal{X}_k} \alpha^2_{xx'y} \log k_{x'}$$

Amenities are additive in a worker-firm fixed effect α^1, and a sum of cross log terms in the number of workers of each type in the workforce $\sum_{x' \in \mathcal{X}_k} \alpha^2_{xx'y} \log k_{x'}$. In the remainder of the paper α^1 (resp. $alpha^2$) will be referred to as the first-order (resp. second order) contribution to amenities. This parametrization offers a flexible way of capturing both the baseline non-wage utility worker x receives from working in firm y, and any additional utilities from working with others, either of the same type ($x = x'$), or of different types $x \neq x'$. These also depend on the type of firm y. Both α^1 and α^2 can be above or below zero, so they may capture utilities or disutilities. The use of the log captures the waning size effects wane as the workforce becomes large:

$$\frac{\partial A_{xyk}}{\partial k_{x'}} = \frac{\alpha^2_{xx'y}}{k_{x'}}$$

meaning the effect of an additional worker of Type x' on the utility of worker x in firm y fades out as the number of Type x' increases.

Workforce-firm production is parametrized as:

$$\Gamma_{yk} = \sum_{x \in \mathcal{X}_k} \gamma^1_{xy} k_x + \sum_{x,x' \in \mathcal{X}_k} \gamma^2_{xx'y} k_x \log k_{x'}.$$

Production is additive in two terms: the first-order productivity coefficients γ^1 capture each type's standalone marginal product, and the second-order coefficients γ^2 which are cross-semi-elasticities and capture pairwise complementarity between worker types. Every $\gamma^2_{xx'y}$ and the symmetric $\gamma^2_{x'xy}$ can be above or below zero, so this parametrization is a flexible way of measuring these interactions. On its own, k_x affects production in firms y only through γ^1_{xy} and γ^2_{xxy}, which together are referred to as 'own' effects, while $\gamma^2_{xx'y}$ for $x' \neq x$ are cross effects. Under this parametrization, marginal productivity of a type x worker in firm y is given by:

$$\frac{\partial \Gamma_{yk}}{\partial k_x} = \gamma^1_{xy} + \sum_{x' \in \mathcal{X}_k} \gamma^2_{xx'y} \log k_{x'} + \gamma^2_{x'xy} \frac{k_{x'}}{k_x}.$$

Augmenting the workforce by an additional worker of type x increases (or decreases) production non-linearly in k_x and $k_{x'}$, and the change in production is highly dependent on the other workers in the workforce. The parametrization also offers insights on the complementarity or substitution patterns between workers of different types through the cross-marginal productivity:

$$\frac{\partial^2 \Gamma_{yk}}{\partial k_x \partial k_{x'}} = \frac{\gamma^2_{xx'y}}{k_{x'}} + \frac{\gamma^2_{x'xy}}{k_x}$$

which shows the parametrization allows for complex patterns of complementarity $\left(\frac{\partial^2 \Gamma_{yk}}{\partial k_x \partial k_{x'}} > 0\right)$ and substitutability between workers of different types $\left(\frac{\partial^2 \Gamma_{yk}}{\partial k_x \partial k_{x'}} < 0\right)$. As k_x and $k_{x'}$ grow, the second derivative fades out, consistent with the idea that changes marginal productivity are lower in big workforces. One can easily show that

$$\frac{\partial^2 \Gamma_{yk}}{\partial k_x \partial k_{x'}} > 0 \Leftrightarrow \begin{cases} \frac{k_x}{k_{x'}} > -\frac{\gamma^2_{xx'y}}{\gamma^2_{x'xy}} \text{ if } \gamma^2_{xx'y} > 0 \\ \frac{k_x}{k_{x'}} < -\frac{\gamma^2_{xx'y}}{\gamma^2_{x'xy}} \text{ if } \gamma^2_{xx'y} < 0 \end{cases}$$

If $\gamma^2_{xx'y} > 0$, x and x' are complements in workforces in which the ratio $\frac{k_x}{k_{x'}}$ is large; that is, there are more workers of type x than x'. In fact, if the symmetric parameter is also above zero - that is $\gamma^2_{x'xy} > 0$ - then the ratio $-\frac{\gamma^2_{xx'y}}{\gamma^2_{x'xy}}$ is smaller than 0, and since k_x and $k_{x'}$ are positive, the condition is always satisfied and x and x' are always complements. On the other hand, if $\gamma^2_{xx'y} < 0$ and $\gamma^2_{x'xy} < 0$, then $-\frac{\gamma^2_{xx'y}}{\gamma^2_{x'xy}}$ is also always negative, and x and x' are always substitutes. Lastly, if $\gamma^2_{xx'y}$ and $\gamma^2_{x'xy}$ are of opposite signs, the pattern depends on the relative number of x and x' workers in the workforce. This is in contrast to a CES functional form, which is commonly chosen to model production functions in the literature. The CES imposes a constant elasticity of substitution between workers of different types, and would also require nesting production by workers' type dimension, which is not necessary here.

Under the functional forms chosen for A_{xyk} and Γ_{yk}, total match surplus Φ_{yk} is

$$\Phi_{yk} = \sum_{x \in \mathcal{X}_k} \phi^1_{xy} k_x + \sum_{x,x' \in \mathcal{X}_k} \phi^2_{xx'y} k_x \log k_{x'}$$

where $\phi^1_{xy} = \alpha^1_{xy} + \gamma^1_{xy}$ and $\phi^2_{xx'y} = \alpha^2_{xx'y} + \gamma^2_{xx'y}$. The surplus function Φ inherits the same properties discussed above for the production function Γ. It relates to the one-to-one matching literature, in which matching is said to be assortative on a characteristic if the surplus function's cross-derivative in the levels of that characteristic on both sides of the market is positive; that is, if both sides are complement in the surplus (Becker (1973), Fox (2010)). The same logic applies in the one-to-many parametrization proposed here, but the inputs are the number of workers instead of the level of a characteristic. The one-to-many surplus Φ features sorting between the firm and the worker, driven by the linear term ϕ^1_{xy}, which is the classic object studied in the one-to-one matching literature, as well as sorting between workers, or worker segregation, through the log linear terms $\phi^2_{xx'y}$ and $\phi^2_{x'xy}$.

Let $\phi = \left(\left(\phi^1_{xy} \right)_{x,y}, \left(\phi^2_{xx'y} \right)_{x,x',y} \right)$ and $\alpha = \left(\left(\alpha^1_{xy} \right)_{x,y}, \left(\alpha^2_{xx'y} \right)_{x,x',y} \right)$ represent the set of total surplus and amenity parameters to estimate. The set of productivity parameters $\gamma = \left(\left(\gamma^1_{xy} \right)_{x,y}, \left(\gamma^2_{xx'y} \right)_{x,x',y} \right)$ can be deduced from estimation of ϕ and α.

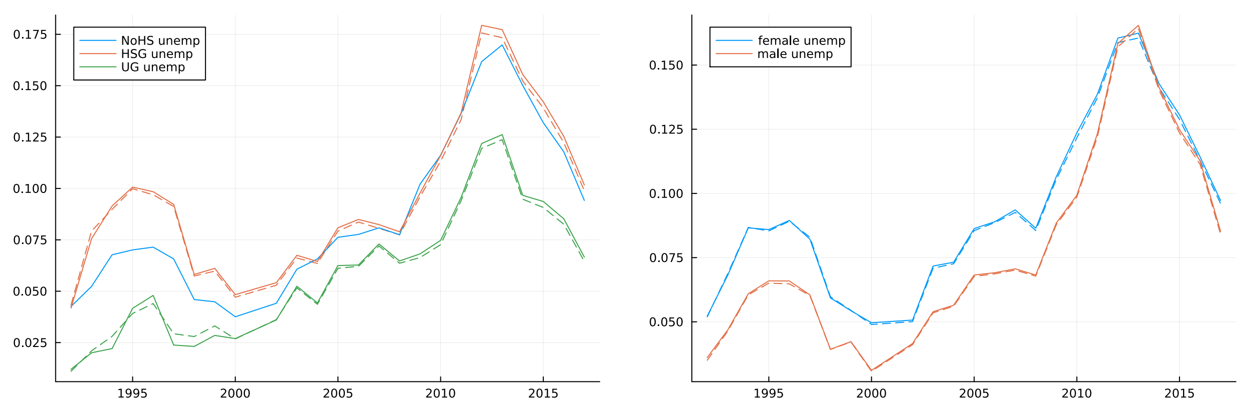
3.3 Generalized Method of Moments

The model is estimated in two steps, where the two sets of parameters ϕ and α are estimated separately. In the first step, total surplus parameters ϕ are estimated by fitting matching moments; and in the second step, amenity parameters α are estimated by fitting wage moments. Estimation requires observing the matching and wage distributions across worker types, firm types and workforce compositions. The model is just-identified, as the number of parameters to estimate is equal to the number of moments used in the estimation. Estimation is run on every year observed in the data separately. Because of its high-dimensionality, the space of workforce compositions $\mathcal{K}$ is reduced before estimation. The full procedure for estimation is described in Appendix C. It obtains estimates $\hat{\phi}$ in the matching step and $\hat{\alpha}$ in the wage step for each year. Productivity parameters $\hat{\gamma}$ are then deduced from the difference between $\hat{\phi}$ and $\hat{\alpha}$. Because the parameters are estimated form observed matching and wages, and not directly inferred from direct observation of productivity and amenities, they should be interpreted as perceived, or apparent productivity and amenities.

3.4 Model Fit

The model is estimated yearly on *Quadros de Pessoal* from 1991 to 2017, using the same grid $\mathcal{K}^c$ in every year, obtaining estimates $\hat{\phi}^t$ and $\hat{\alpha}^t$ for $t \in \{1991, \dots, 2017\}$. The resulting fit on matching and wage is satisfactory, the following figures show. Figure 1 plots the observed and predicted unemployed workers' shares by education level and gender. Both the respective levels of unemployment by education level and gender and the unemployment peaks in 2012 are matched by the model.

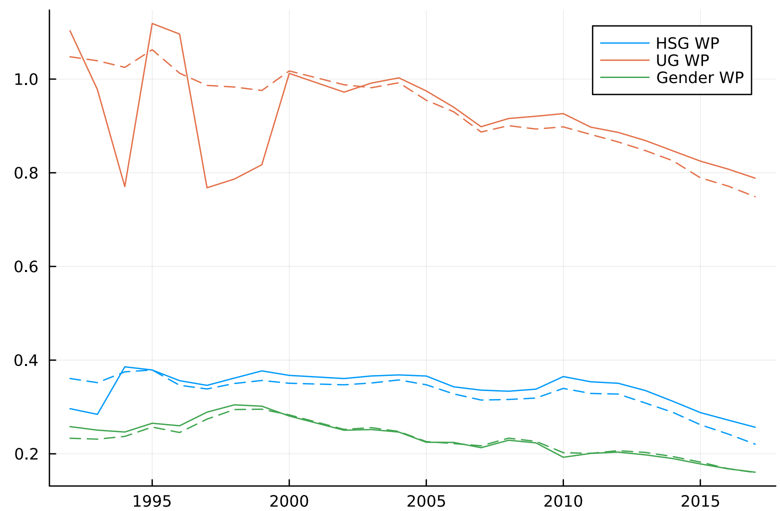
Figure 1: Model Fit - Unemployment



Source: *Quadros de Pessoal*. Author's own calculations from model's estimation. The dotted lines show the observed unemployment shares $\frac{\tilde{\mu}_{x0}}{\sum_{y,k\in\tilde{\mathcal{K}}} k_x \tilde{\mu}_{yk} + \tilde{\mu}_{x0}}$ where x is an education level on the left and a gender on the right. The solid lines show the predicted unemployment shares $\frac{\hat{\mu}_{x0}}{\sum_{y,k\in\mathcal{K}^c} k_x \hat{\mu}_{yk} + \hat{\mu}_{x0}}$ after model estimation.

Figure 2 plots the observed and predicted wage gaps by education and gender. The observed university wage premium, i.e. the average wage gap between university graduates and workers with no degrees is clearly declining over time, as captured by the model: it goes from 1.1 in 1991 (i.e. university graduates earned 110% more than less-educated workers on average) to abouxt .8. The model also captures the slower decline in the high school wage premium and the gender wage gap. It underestimates the university wage gap in 1994 and 1997 but does capture the trend over the long-term.

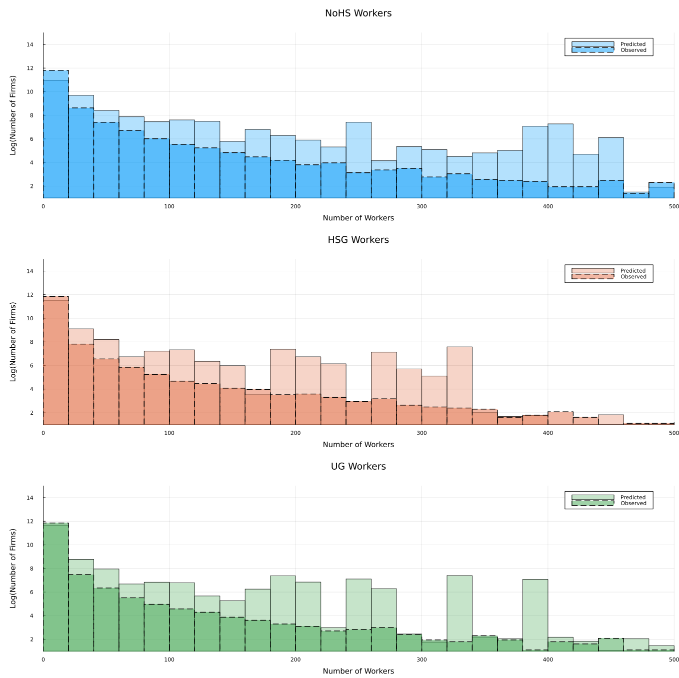
Figure 2: Model Fit - Wage Gaps



Source: *Quadros de Pessoal*. Author's own calculations from model's estimation. The dotted lines show the observed wage gaps $\log(\overline{\tilde{w}}_x) - \log(\overline{\tilde{w}}_{x'})$ where $\overline{\tilde{w}}_x$ is the average wage by either education levels or gender. The solid lines show the predicted wage gaps $\log(\overline{\hat{w}}_x) - \log(\overline{\hat{w}}_{x'})$ where $\overline{\hat{w}}_x = \frac{\sum_{y,k} \hat{\mu}_{yk} \hat{w}_{xyk}}{\sum_{y,k} \mu_{yk}}$

Finally, Figure 3 plots the observed and predicted log count of firms by the number of workers of each education level that they employ. This moment is not targeted in the estimation, but the model captures the distribution of firms by workforce composition fairly well: many firms employ few workers of each education level, which is matched by the predicted distribution. The model also captures the long right tail of the distribution, although it does overestimate the number of large workforces.

Figure 3: Model Fit - Workforce Composition



Source: *Quadros de Pessoal*. Author's own calculations from model's estimation. The dotted bars show the observed log count of firms by number of workers of a given education level that are employed by the firm. The predicted bars show the predicted log count. The x-axis is capped at 500 workers for readability. Both the observed and predicted counts are very small after 500 workers.

4 Results

The goal of the model estimation is to reveal the supply and demand forces that shape the sorting patterns and wage gaps in the Portuguese labor market. This section begins by outlining the empirical facts that motivate the estimation. It then describes the firm classification results and presents the estimated parameters.

4.1 Context

The Portuguese labor market has undergone significant changes over the past three decades. On the demand side, the country entered the European Union in 1986, which fuelled its economy's transition from being dominated by manufacturing (45% of the labor force employed in 1991) to services (32% of the labor force employed in 2017). Table 2 shows how the education expansion and the feminization of the labor market are reflected in the sample. Together, university and high school graduates account for only 14.3% of the labor force in 1991. By 2017, this share had more than tripled to 50.1%. The secondary education system was reformed once in 1986, and then again in 2009. In 1986, Portugal uniformized its primary and secondary school system to meet the European Economic Community requirements. New schools were also built throughout the 1990s in rural areas. The 2009 reform made secondary education compulsory, by requiring students to stay in school until 18 years old. Consequently, the largest increases in the share of high school graduates occurred in the 1990s and again in the 2010s. The late 1990s and early 2000s saw an enrollment boom in public universities and polytechnic institutes, which explains the increase in the share of university graduates in that period. Both public universities and polytechnic institutes are virtually free (tuition fees are only around a few hundred euros a year). The former train students for academic degrees, while the latter offer vocational, technical and applied training. Graduates from both types of institutions are bundled as 'university' graduates throughout the rest of the paper. As a point of comparison, the percentage of high school graduates in the US workforce increased from 75% to 90% over the same period[8]. The high school wage premium stagnated around 35% in the 1990s, and declined to 22% in the 2010s. The university wage premium declined gradually in the 1990s and early 2010s, and also dropped in the late 2010s from 87% to 75%.

Women comprised approximately one-third of the sample in 1991, and 45% in 2017. Their entry into the labor market was likely fueled by their growing education rates: they accounted for about one-quarter of university graduates in the labor force in 1991, and more than half in 2017. The gender pay gap increased in the 1990s to reach 27% in 1997, and then declined to 16% in 2017.

Table 2: The Portuguese Labor Market over Time

	1991	1997	2002	2007	2012	2017
% University Grad.	3.9	2.7	9.1	13.4	17.0	20.8
% High school Grad.	10.4	17.1	18.6	22.0	24.9	29.3
University WP	1.05	.99	.99	.89	.87	.75
High school WP	.36	.34	.35	.31	.33	.22
% Women	30.6	38.1	40.4	41.4	45.6	45.5
% Women among Uni Grads.	25.2	36.2	46.4	51.8	55.3	56.6
Gender wage gap	.23	.27	.25	.22	.21	.16

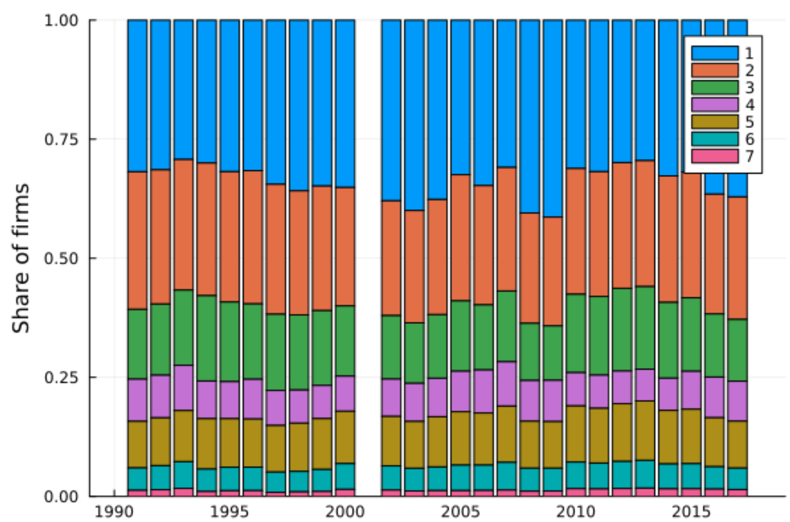
Source: *Quadros de Pessoal.* Author's own calculations. Shares are computed from the active population (employed and unemployed). High school graduates are exclusive of university graduates. WP (wage gaps) are computed as the difference of average log wage by education level, with respect to the default less-educated workers (who did not complete high school).

[8]Percentages computed over workers aged more than 25, Census data

4.2 Firm Classification

The first step in the estimation strategy is to classify firms into types based on their wage distribution, using the method described in section 3.2. The resulting classification is then used in the model's estimation as firm types. Figure 4 plots the distribution of firm types over the years. Low type firms (Types 1,2,and 3) comprise approximately 75% of the firms in the data, which shows there is a high share of relatively low-paying firms with similar wage distribution throughout the 1990s until 2017. Note that this classification strategy allows firms to change type over time, depending on changes in their wage distribution[9]. These changes are consistent with the assumption that the wage distribution reflects the firms strategic decisions (management quality, capital investments, etc.) which may evolve over time. In the aggregate, the distribution of firm types is stable over time.

Figure 4: Firm types distribution over time



Source: *Quadros de Pessoal*. Author's own calculations based on the clustering of within firm wage distribution.

Table 3 presents some descriptive statistics by firm type. Average log wage is increasing in firm type by construction. The average number of employees per firm is not monotonous in firm type, although high-type firms tend to employ more workers than low-type firms. The share of high school graduates (HSG) is bell-shaped in firm types: the types that employ the most HSGs are types 3 to 6. Overall, every firm type employ at least a quarter of HSGs on average. The share of university graduates (UG) varies much more across firm types and increases monotonically with firm type. High types firms' workforces are composed of more than 50% of UGs on average. The share of women declines with firm type, from 46% in

[9] On average 29.1% of firms change type every year

Type 1 firms to 39% in Type 7 firms. The share of firms in manufacturing also tends to decrease with firm type, while the share of firms in services increases. One exception to this rule is Type 4 that has a higher share of manufacturing firms and a lower share of service firms than Type 3 and 5.

Table 3 shows the *kmeans* group classification method captures substantial differences in productivity between firms, but offers no explanation on the mechanisms that drive these differences. By allowing for different worker-firm sorting and worker interdependence patterns by firm type, the model estimation presented next sheds light on the drivers of these differences.

Table 3: Firm Types in 2017 - Descriptive Statistics

Firm type	1	2	3	4	5	6	7
Average log wage	1.34	1.5	1.69	1.7	1.94	2.28	2.82
Average nb. of employees	6.7	13.8	9.3	37.4	23.4	33.1	58.1
Sh. HSG	0.24	0.29	0.31	0.3	0.31	0.3	0.26
Sh. UG	0.06	0.13	0.22	0.21	0.36	0.51	0.65
Sh. Women	0.46	0.43	0.4	0.46	0.43	0.43	0.39
Share Manufacturing	0.19	0.17	0.13	0.22	0.14	0.08	0.05
Share Services	0.18	0.27	0.34	0.3	0.34	0.4	0.45

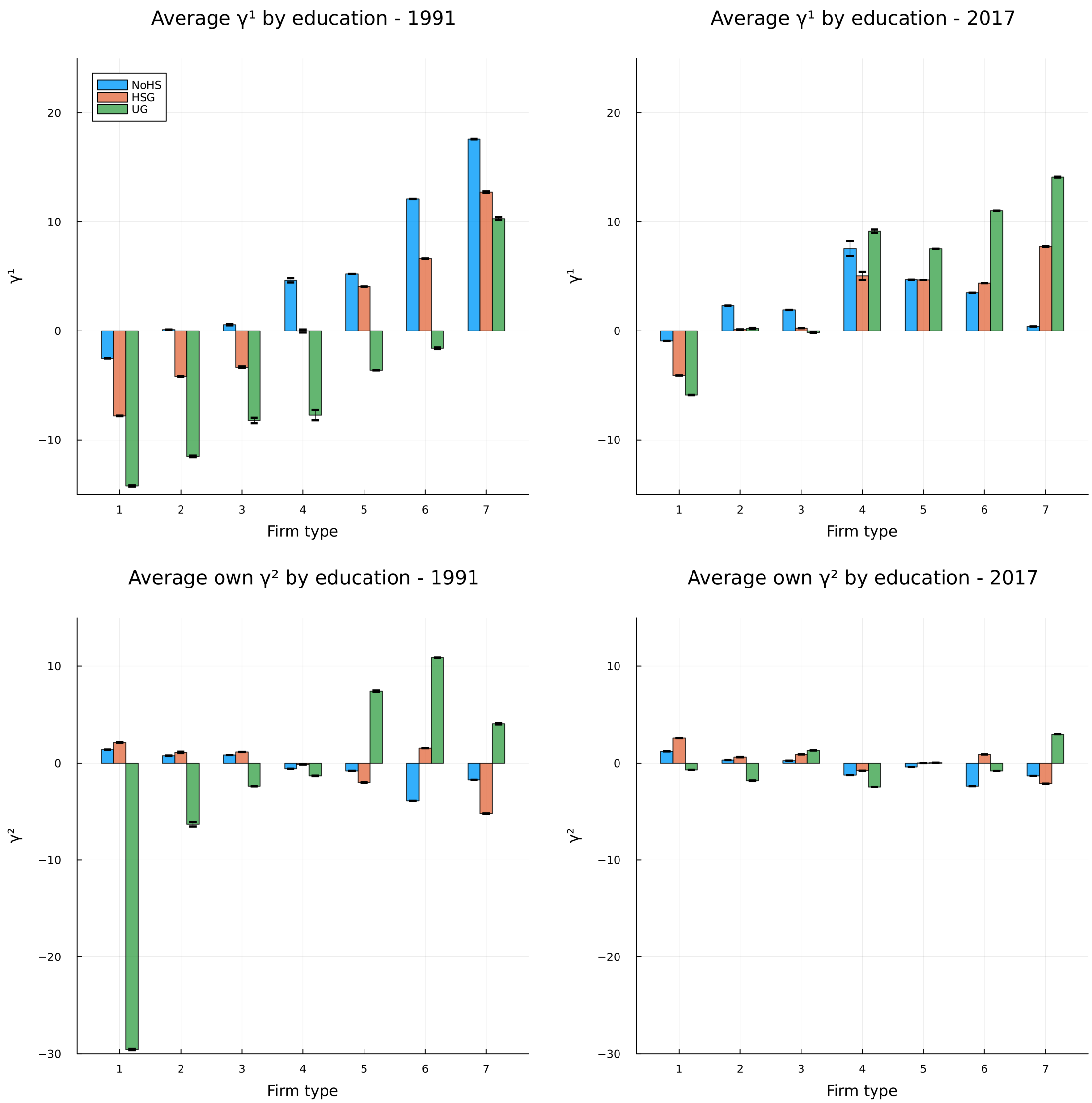
Source: *Quadros de Pessoal*. Author's own calculations. log wage is is log euros/hours. All shares and averages are first computed at the firm level, then averaged out across firm types. HSG: high school graduates, UG: university graduates. Sectors that are neither manufacturing nor services are: agriculture, fishing, hunting, extractive industries, and construction - trade and hospitality - transport and commmunications.

4.3 Estimated Parameters

Given estimated firm types, the sets of productivity parameters $\gamma = \left(\left(\gamma^1_{xy}\right)_{x,y}, \left(\gamma^2_{xx'y}\right)_{x,x',y}\right)$ and amenity parameters $\alpha = \left(\left(\alpha^1_{xy}\right)_{x,y}, \left(\alpha^2_{xx'y}\right)_{x,x',y}\right)$ are estimated every year using the method described in section 3. As noted in the discussion of Proposition 3, the interpretation of amenities and production parameters for a give worker type, firm type and workforce is relative to the other worker-firm-workforces combinations, or to the worker's outside option. The estimated parameters presented here are therefore understood in comparison to each other, and carry no absolute meaning. This sub-section focuses on the parameters most relevant to the change in the education wage premium, while Appendix D presents the full set of estimated parameters.

Figure 5 shows the average first- and second-order own- contribution to production by education level and firm type, in 1991 and 2017. The averages are weighted by the mass of workers of each type in the labor market. Both productivity and amenity are in euros per hour. The second order contribution are own contribution, i.e. they refer to the parameters γ^2_{xxy}. Cross-contributions are presented in the appendix D. Average γ^1 vary widely by education levels and clusters, both in 1991 and 2017. In both 1991 and 2017, the higher the firm type the higher the productivity contribution, across all education levels, which confirms high-type firms are more productive. In 1991, university graduates (UG) contribution is very low in low-type firms compared to high-type firms. Less-educated workers' (NoHS) linear productivity is higher than other education levels in all firm types. High school graduates (HSG) is slightly higher than university graduates' in all firm types. These observations suggest the Portuguese economy in 1991 relied very little on skilled labor. In 2017, average linear contributions to productivity are less dispersed than in 1991. University graduates have overtaken high school graduates and less-educated workers in medium and high firm types. In low-type firms, the gap between average linear productivities across education levels has shrunk. We can see that skill-biased technological change happens between 1991 and 2017 in Portugal, but it is uneven across firm types: the lowest firm types experience stronger increases in productivity for university and high school graduates than the highest firm types. High school graduates even suffer from a decrease in productivity compared to university graduates in high-type firms. Average own γ^2 are also more dispersed in 1991 than they are in 2017, especially for university graduates. In 1991, their second-order contribution to productivity is much lower than other education levels in low type-firms and much higher in high type-firms. There are therefore large returns to productivity from hiring large numbers of university graduates in high-type firms, which compensates the low linear contributions observed for this type of workers. The differences between low and high-type firms and education levels are much more narrow in 2017. In 2017, less-educated workers' second-order contributions to productivity are of opposite sign to university graduates': there are positive in low-type firms and negative in high-type firms, causing the polarized sorting into firms by education level (less-educated workers employed in low-type firms, university graduates in high type firms). The narrowing of the gap between low and high-type firms for university graduates between 1991 and 2017 can also be interpreted as a form of uneven technological change: the strength of complementarity of the most educated workers increases in low-type firms, but it reduces in high-type firms.

Figure 5: Own Contributions to Productivity - Education



Source: *Quadros de Pessoal*. Author's own calculations from estimated parameters. Weighted averages by education are computed over both gender, e.g. for a given firm cluster y and education level NoHS: $\gamma^1_{\text{NoHS},y} = \frac{m_{\text{F-NoHS}}\gamma^1_{\text{F-NoHS},y} + m_{\text{M-NoHS}}\gamma^1_{\text{M-NoHS},y}}{m_{\text{F-NoHS}} + m_{\text{M-NoHS}}}$ and $\gamma^2_{\text{NoHS},textNoHS,y} = \frac{m_{\text{F-NoHS}}\gamma^1_{\text{F-NoHS},textNoHS,y} + m_{\text{M-NoHS}}\gamma^1_{\text{M-NoHS,M-NoHS},y}}{m_{\text{F-NoHS}} + m_{\text{M-NoHS}}}$. Euros per hours. Confidence intervals are provided at the 95% level.

Appendix D presents average own contributions to amenities by education level (Figure 6), to productivity and amenities by gender (Figures 7, 8), and average cross-contributions by education level and gender (Figures 10, 9). These figures offer three main takeaways. First, cross-complementarities in production between university graduates decrease for men and increase for women between 1991 and 2017. The effect on university graduates' average wage is therefore ambiguous. Second, on the amenities side, first-order parameters tend to equalize over time across education levels, which all other things equal also equalizes wages. Third, both productivity and amenities parameters by gender tend to converge over time, which works to reduce the gender pay gap. The conclusions from estimation regarding the education wage premia are ambiguous: university graduates benefit from an increase in productivity compared to less-educated workers, but mostly in low-type firms, while high school graduates suffer from a relative decrease in productivity in high-type firms. Because low-type firms remain less productive than high-type firms, the uneven skill-biased technological change modifies sorting patterns of educated workers, but does not necessarily increase the education wage premium.

The estimated values for pairwise complementarity parameters γ^2 and α^2 are presented in Figures 9 and 10 in Appendix D, along with a detailed analysis. This section focuses on the complementarity and substitutability patterns in production between worker types: if $\frac{\partial^2 \Gamma_{yk}}{\partial k_{x'} \partial k_x} > 0$ for all k, workers x and x' are said to be complements for firm type y. If $\frac{\partial^2 \Gamma_{yk}}{\partial k_{x'} \partial k_x} < 0$ for all k, they are substitutes. In many cases, whether the second derivative is above or below zero depends on the workforce k, as described in subsection 3.2. Table 4 provides a general overview of the substitutability and complementarity patterns between worker types in 1991 and 2017. 40% of worker pairs are neither substitutes nor complements in 1991 (35% in 2017), which means that whether they are substitutes or complements depends on the workforce. 31% of worker pairs are always substitutes in 1991 (36% in 2017), and 29% are always complements in 1991 (29% in 2017), regardless of the relative number of Type x and x' workers in the workforce. The nature of the interdependence between workers when a pair is indeterminate depends on the ratio $-\gamma^2_{xx'y}/\gamma^2_{x'xy}$ and the sign of $\gamma^2_{xx'y}$. The last two columns in Table 4 show average ratios in each case. If $\gamma^2_{xx'y} > 0$, a ratio above 1 means x and x' are substitutes if $k_x < k_{x'}$, and complements otherwise. If $\gamma^2_{xx'y} < 0$, a ratio above 1 means x and x' are complements if $k_x < k_{x'}$, and substitutes otherwise. The average ratios change between 1991 and 2017. In 1991 they are large, at 10.8 and 7.8, meaning there should be a large gap in numbers between Type x and Type x' workers for the pattern to switch. In 2017 they were not far from one, which suggests the interdependence may switch between complementary or substitute worker pairs for similar workforces, if one or two additional workers were added or removed to the workforce. These results show that choosing a functional form that restricts worker pairs to be always substitutes or complements, such as the constant elasticity of substitution is restrictive in this setting.

Table 4: Substitutability and Complementarity Patterns, 1991 vs. 2017

	% Worker Pairs			Mean $-\gamma^2_{xx'y}/\gamma^2_{x'xy}$ (indeterminate pairs only)	
	Substitute	Complement	Indeterminate	$\gamma^2_{xx'y} > 0$	$\gamma^2_{xx'y} < 0$
1991	31	29	40	10.8	7.8
2017	36	29	35	2.1	1.3

Source: *Quadros de Pessoal.* Author's calculations from estimated parameters. Each year covers $7 \times 21 = 147$ ordered worker-type pairs. The mean ratio in the last two columns is calculated only for those pairs whose interaction is classified as indeterminate.

5 Counterfactuals

The main question in this paper concerns how supply and demand have evolved over time. The first three counterfactuals aim to isolate the impact of these changes on the education wage premia. The results show that the Portuguese education expansion exerts downward pressure on the both the high school and university wage premia, as is expected from an increase in relative supply of educated workers. On the demand side, two types of productivity change are studied: the change in own productivity (from a single's worker type), and the change in cross-productivity (from interdependence between worker types). The changes in own and cross-productivity over time have opposing effects on the education wage premium: changes in the linear own-productivity favor educated workers, while changes in linear-log own- and cross-productivity push the university wage premium downward but the high school wage premium upward. The last counterfactual presented in this section examines the impact of differential cross-productivity and cross-preferences across genders on the gender pay gap. It finds the high complementarities in production between men are an important driver of the gender pay gap. Finally, the counterfactual in the appendix D investigates how workers' different preferences across firm types shift education wage premia.

5.1 Supply and Demand changes

5.1.1 Education shares fixed to 1991

The first counterfactual simulates a world in which the education expansion has not happened, while all other economic and demographic changes occurred as observed or estimated. To do so, it fixes the shares of individuals with no degree, high school and university graduates to their 1991 values. These shares are computed within gender. All other parameters (productivity, amenities) vary as estimated and primitives (total mass of active individuals, gender shares, firm masses) vary as observed. Table 5 shows sorting of university and high school graduates in high and medium type firms (Types 6-7 and 4-5, respectively). Between 1991 and 2017, the predicted shares of university graduates employed in high-type firms decreases by 11.6. (percentage points), while sorting into medium type firms increases by about 6.5. In the counterfactuals, sorting into medium type firms increases by approximately a similar amount, but sorting into high type firms drops by 35.6. Similar trends are observed for high school graduates. In the counterfactual world, where the education expansion does not happen, sorting of educated workers into low-type firm grows at the expense of sorting into high-type firms. This is caused by the increase in productivity of low type firms, which pulls the limited numbers of educated workers from high-type firms.

Table 5: Fixed educ. shares - sorting

	% UG in				% HSG in			
	high types		med. types		high types		med. types	
	1991	2017	1991	2017	1991	2017	1991	2017
Predicted	48.2	36.6	35.8	42.3	37.6	17.1	38.4	38.7
Count. educ. sh.	48.2	12.6	35.8	43.8	37.6	5.3	38.4	32.6

Source: *Quadros de Pessoal*. Author's calculations from model estimates. High types: Types 6-7, med types: types 4-5. UG: university graduates. HSG: high school graduates. Shares are computed among all employed workers of a given education level.

Table 6 shows the predicted and counterfactual wage gaps. Its main takeaway is that, absent the education expansion the education wage gaps would have increased, instead of decreasing as in the baseline. The university wage gap would have risen by 21., and the high school wage gap by 40. between 1991 and 2017. Although educated workers sort into low-type firms in this counterfactual, their increased productivity combined with their scarcity allows them to increase their wage in these firms.

Table 6: Fixed educ. shares - wage gaps

	UG-NoHS wage gap			HSG-NoHS wage gap		
	1991	2017	Δ	1991	2017	Δ
Predicted	1.1	.79	-.33	.28	.26	-.02
Count. educ sh.	1.1	1.3	.21	.28	.67	.40

Source: *Quadros de Pessoal*. Author's calculations from model estimates. UG: university graduates. HSG: high school graduates. NoHS: less educated workers. Wage gaps are computed as the difference of the log of the average wage of a given education level and the log of the average wage of less-educated workers (without a high school degree, NoHS). The average wage is computed over all employed workers of a given education level.

5.1.2 Productivity parameters fixed to 1991

This section runs three counterfactuals. In the first counterfactual, first-degree parameters γ^1 are fixed to their 1991 values throughout the time period. All other model primitives (worker and firm masses) and parameters (amenities α^1 and α^2, pairwise complementarity in productivity γ^2) are allowed to vary as observed or estimated. In the second counterfactual, own-productivity parameters are fixed to their 1991 estimated values. These include all of first-order productivity parameters γ^1, and second-order parameters γ^2_{xxy} for all x and y. In the third counterfactual, the interdependence parameters $\gamma^2_{xx'y}$ for all x, x',y such that $x' \neq x$ are fixed to their 1991 estimated values, while all other primitives and parameters vary. Tables 7 presents the predicted and counterfactual shares of university (UG) and high school graduates (HSG) employed in high-type firms (types 6 and 7). Since the parameters are fixed to their 1991 value, the predicted and counterfactual sorting in 1991 is the same. In 2017, the majority of university and high school graduates sort in high-type firms in the counterfactuals. As shown in subsection 4.3, educated workers are very productive in high-type firms and very unproductive in low-type firms in 1991. Between 1991 and 2017, their productivity increases mostly in low-type firms. This increase does not occur in the counterfactual, which polarizes the sorting of educated workers in high-type firms. The polarization is more pronounced for university than for high school graduates. All three counterfactuals display the same polarization so both the own-productivity and cross-productivity parameters contribute.

Table 7: Fixed own- and cross-productivity - sorting

	% UG in high firm types		% HSG in high firm types	
	1991	2017	1991	2017
Predicted	48.2	36.6	37.6	17.1
Count. first-deg. prod	48.2	94.5	37.6	80.4
Count. own prod.	48.2	97.7	37.6	96.9
Count. cross prod.	48.2	80.0	37.6	77.6

Source: *Quadros de Pessoal*. Author's calculations from model parameters. High types: Types 6-7. UG: university graduates. HSG: high school graduates. Shares are computed among all employed workers of a given education level.

Table 8 presents the predicted and counterfactual education wage gaps, comparing wages of university and high school graduates to less educated workers. Both education wage gaps are larger in the first counterfactual compared to the predicted baseline in 2017, to -.11 for university graduates and -.41 for high school graduates. This is due to the relative increase in first-order productivity terms for educated workers: had it not happened as in the counterfactual, they would have been worse off. Absent the increase in productivity of educated workers in low-type firms, they sort more into high-type firms, but the polarized sorting is not enough to offset the lack of productivity gain. Therefore, these gains in educated workers' productivity, also referred to as skill-biased technological change in the literature, are a key driver of the education wage premium in Portugal between 1991 and 2017. In the second and third counterfactuals, the university wage gap moves in the opposite direction: it increases to 1.67 and 1.07, respectively. This means that the actual change in the second degree parameters γ^2, and especially own-productivity parameters γ^2_{xxy} brings the university wage premium down, by reducing complementarities for university educated workers in high type firms. Indeed, the estimates presented in Figure 5 and Appendix D show the second-degree parameters γ^2 shift from being overall very positive in 1991 to a much smaller range in 2017 for university educated workers. Technological change therefore has two counteracting effects on the university wage premium. The high school wage premium offers the opposite picture: it is lower in all counterfactuals than in the predicted baseline. Technological change is therefore unambiguous for high school graduates: both own and cross-production parameters evolve in their favor.

Table 8: Fixed own- and cross-productivity - wage gaps

	UG-NoHS wage gap			HSG-NoHS wage gap		
	1991	2017	Δ	1991	2017	Δ
Predicted	1.12	.79	-.33	.28	.26	-.02
Count. first-deg. prod	1.12	-.11	-1.23	.28	-.41	.68
Count. own prod.	1.12	1.67	.55	.28	.09	-.19
Count. cross prod.	1.12	1.07	-.04	.28	-.49	-.77

Source: *Quadros de Pessoal*. Author's calculations from model estimates. UG: university graduates. HSG: high school graduates. Wage gaps are computed as the difference of the log of the average wage of a given education level and the log of the average wage of less-educated workers (without a high school degree, NoHS). The average wage is computed over all employed workers of a given education level.

5.2 The Gender Pay Gap

The fourth counterfactual focuses on the second-degree productivity parameters γ^2, on the gender dimension. The goal is to assess whether interdependence in production between and within genders affect the gender pay gap. The estimates for γ^2 presented in Appendix D (Figures 10, 9) show there are large differences between women and men university graduates. In high type firms, the interdependence between men in production is strongly positive, and it is strongly negative between women. The opposite is true with regard to amenities: male university graduates strongly dislike working with one another. Together these two findings strengthen the gender pay gap by sorting male university graduates in highly productive firms, while compensating them for their dislike of each other in these firms. In the counterfactual, all second-degree parameters $\gamma^2_{xx'y}$ where either x or x' refers to a female worker, are set to their counterpart where x or x' refers to a male worker. In other words, interdependence in production depends only on education levels and firm types but is equal across genders. The parameters are allowed to change over time as estimated. The effect on sorting for both gender is strong, as shown in Table (9). In the counterfactual, all women and all men sort in high-type firms in 1991, due to men's positive interdependence being also applied to women. These effects wear off slightly in 2017, as complementarity effects are smaller in high type firms. Men in 2017 tend to sort much less in medium-type firms than in the predicted baseline, reflecting their lower preference for amenities in these firms.

Table 9: Equalized cross-productivity across genders - sorting

	% Women in				% Men in			
	high types		med. types		high types		med. types	
	1991	2017	1991	2017	1991	2017	1991	2017
Predicted	15.4	13.5	32.2	36.2	19.7	17.1	39.5	35.9
Count. equal cross prod	100	88.6	2.4	32.8	100	88.3	33.2	2.0

Source: *Quadros de Pessoal*. Author's calculations from model estimates. High types: Types 6-7, med types: types 4-5. Shares are computed among all employed workers of a given gender.

Table 10 shows drastic differences in the gender pay gap between the predicted baseline and the counterfactual. In the baseline, the gender pay gap decreased from .26 to .16 between 1991 and 2017. In the counterfactual, the intensive sorting of women into high-type firms results in a much smaller gender wage gap in 1991, at only .06. This finding strongly suggests, much of men's higher average pay stems from their strong complementarity in production, rather than first-degree contributions or their dislike for each other evidenced in the amenities. In 2017, the gender pay gap is inverted: women earn about 25% more than men. This is due to the relative productivity of women in high and medium type firms increasing between 1991 and 2017, while in the counterfactual the interdependency is kept the same as men's. These striking results show pairwise complementarity in perceived productivity, rather than own contribution to firm production, plays a key role in the gender pay differences in Portugal.

Table 10: *Equalized cross-productivity across genders - wage gaps*

	Gender wage gap		
	1991	2017	Δ
Predicted	.26	.16	-.10
Count. equal cross prod	.06	-.25	-.48

Source: *Quadros de Pessoal.* Author's calculations from model estimates. Wage gaps are computed as the difference of the log of mens' average wage and the log of women's average wage. The average wage is computed over all employed workers of a gender.

6 Conclusion

The paper proposes a novel model of one-to-many matching, and uses it to study the compression of the wage structure in the Portuguese labor market along the education and gender dimensions. The model generalizes the one-to-one random utility matching framework to a one-to-many setting, and enables the estimation of workers non-wage amenities and firms' productivity using observed matching and wage distributions. Crucially, the one-to-many framework allows identification of cross-complementarity terms between workers in both non-wage amenities and productivity, in addition to the standard complementarity between workers and firms. This makes it possible to identify how worker segregation impacts wage gaps, along with worker and firm sorting.

The wage structure in Portugal has compressed over the last 35 years along the education and gender dimensions. The model estimation and counterfactuals show that the compression is driven by a combination of factors. First, the Portuguese education expansion has had the largest negative impact on the university wage premium; the premium would have been 54p.p. higher in 2017 in the absence of the expansion. The expansion has also negatively impacted the high school wage premium (by 42p.p.), but the largest negative driver of this wage premium is actually workers' differential non-wage preferences across firms (by 70p.p.). Educated workers' productivity increased over the period, placing upward pressure on the education wage premia, but it has done so non-uniformly across firms: low type firms experience more technological change than high type firms. Finally, complementarities in productivity matter enormously for the gender pay gap: if women benefitted from the same positive cross-complementarities as men in high type firms, the gender pay gap would be greatly reduced. Changes in cross-complementarities in productivity between education levels also matter for the education wage premia, but differently to university and high school: they drive the former down and the latter up.

This paper offers a novel perspective on the reasons behind changes in the wage structure, by focusing on the role of non-wage amenities and workers cross-complementaries, along with the more standard productivity and firm-worker sorting channels. It also contributes to the growing literature on the compression of the wage structure along the education dimension, which has been shown to occur in countries other that Portugal, such as the United States. Future avenues for research include leveraging the model's matching and wage predictions by workforce composition for estimation and accounting for workers' heterogeneity within education levels and gender.

A Model Appendix

Proof. Proposition 1

Supply and demand are defined as the probabilities that a given option is preferred to all others. The proof consists in deriving a closed form solution for these probabilities under Assumption 1.

$$
\begin{aligned}
&\mathbb{P}\left[(\bar{y},\bar{k}) = \underset{y\in\mathcal{Y},k\in\mathcal{K}}{\arg\max}\left\{u_{xyk}+\varepsilon_{yk},\varepsilon_0\right\}\right]\\
&=\mathbb{P}\left[u_{x\bar{y}\bar{k}}+\varepsilon_{\bar{y}\bar{k}} \geq u_{xyk}+\varepsilon_{yk}\,\forall y,k\,, u_{x\bar{y}\bar{k}}+\varepsilon_{\bar{y}\bar{k}}\geq \varepsilon_0\right]\\
&=\mathbb{P}\left[\varepsilon_{yk}\leq u_{x\bar{y}\bar{k}}-u_{xyk}+\varepsilon_{\bar{y}\bar{k}}\,\forall y,k\,,\varepsilon_0\leq u_{x\bar{y}\bar{k}}+\varepsilon_{\bar{y}\bar{k}}\right]\\
&=\int_{-\infty}^{\infty}\exp\left(-\exp(-u_{x\bar{y}\bar{k}}-\varepsilon)\right)\prod_{y,k\neq\bar{y},\bar{k}}\exp\left(-\exp(-u_{x\bar{y}\bar{k}}+u_{xyk}-\varepsilon)\right)f(\epsilon)\,\mathrm{d}\epsilon\\
&=\int_{-\infty}^{\infty}\exp\left(-\exp(-u_{x\bar{y}\bar{k}})\exp(-\varepsilon)\left(1+\sum_{y,k\neq\bar{y},\bar{k}}\exp(u_{xyk})\right)\right)\exp(-\exp(-\varepsilon))\,\mathrm{d}\varepsilon\\
&=\int_{0}^{\infty}\exp\left(-z\left(\exp(-u_{x\bar{y}\bar{k}})\left(1+\sum_{y,k\neq\bar{y},\bar{k}}\exp(u_{xyk})\right)+1\right)\right)\mathrm{d}z\\
&=\frac{1}{\exp(-u_{x\bar{y}\bar{k}})\left(1+\sum_{y,k\neq\bar{y},\bar{k}}\exp(u_{xyk})\right)+1}\\
&=\frac{\exp(u_{x\bar{y}\bar{k}})}{1+\sum_{y\in\mathcal{Y},k\in\mathcal{K}}\exp(u_{xyk})}
\end{aligned}
$$

where $f(\varepsilon) = \exp(-\exp(-\varepsilon))$ is the Gumbel's pdf, and the transition between rows four and five is through change of variable $z = \exp(-\varepsilon)$. A similar reasoning applies to μ_{x0} and D_k^y. □

Proof. Theorem 1 This proof is based on the same proof techniques as Gretsky et al. (1992)

and Galichon and Salanié (2022).

Consider the individual worker i's utility maximization, and the individual firm j profit maximization. By a standard argument from Gretsky et al. (1992), solving these for all workers and firms is equivalent to the following problem over the sum of worker welfare $\sum_i u_i$ and firm welfare $\sum_j v_j$:

$$\begin{aligned} &\inf_{u,v} \sum_i u_i + \sum_j v_j \\ &\text{s.t} \sum_x \sum_{i:x_i=x}^{k_x} u_i + v_j \geq \Phi_{y_j k} + \sum_x \sum_{i:x_i=x}^{k_x} \epsilon_{iy_j k} + \eta_{jk} \quad \forall k, j \\ &\quad u_i \geq \epsilon_{i0} \end{aligned} \tag{11}$$

Take any two u, v such that $\sum_x k_x u_{xyk} + v_{yk} \geq \Phi_{yk}$ and $u_{x0} = 0$ and define

$$\begin{cases} u_i = \max_{y,k}\{u_{x_i yk} + \epsilon_{iyk}\} \\ v_j = \max_k\{v_{y_j k} + \eta_{jk}\} \end{cases}$$

Then (u, v) satisfies (11)'s constraints.

Reciprocally, fix any u_i, v_j that satisfy the constraints in this problem and define

Let

$$\begin{cases} u_{xyk} = \min_{i,x_i=x}\{u_i - \epsilon_{iyk}\} \text{ and } u_{x0} = 0 \\ v_{yk} = \min_{j,y_j=y}\{v_j - \eta_{jk}\} \end{cases}$$

Then the constraint in problem (11) becomes $\sum_x k_x u_{xyk} + v_{yk} \geq \Phi_{yk}$.

Applying the law of large numbers, we get that (11) is equivalent to

$$\begin{aligned} &\min_{u,v} \sum_x n_x G_x(u_x) + \sum_y m_y H_y(v_y) \\ &\text{s.t} \sum_x k_x u_{xyk} + v_{yk} = \Phi_{yk} \quad \forall k, y \\ &\quad u_{x0} = 0 \end{aligned} \tag{12}$$

Rewrite problem (12) as saddle-point:

$$
\begin{aligned}
&\min_{u,v}\max_{\mu}\sum_x n_x G_x(u_x)+\sum_y m_y H_y(v_y)\\
&\quad+\sum_{y,k}\mu_{yk}\left(\Phi_{yk}-\sum_x k_x u_{xyk}-v_{yk}\right)+\sum_x S_0^x(-u_{x0})\\
&=\max_{\mu}\sum_{y,k}\mu_{yk}\Phi_{ky}\\
&\quad-\sum_x n_x \max_u\left\{\sum_y \frac{k_x\mu_{yk}}{n_x}u_{xyk}+\frac{\mu_{x0}}{n_x}u_{x0}-G_x(u)\right\}\\
&\quad-\sum_y m_y \max_v\left\{\sum_y \frac{\mu_{yk}}{m_y}v_{yk}-H_y(v)\right\}\\
&=\max_{\mu}\sum_{y,k}\mu_{yk}\Phi_{ky}\\
&\quad-\left(\sum_x\sum_{y,k}k_x\mu_{yk}\log\frac{k_x\mu_{yk}}{n_x}-\sum_x S_0^x\log\frac{S_0^x}{n_x}-\sum_{y,k}\mu_{yk}\log\frac{\mu_{yk}}{m_y}\right)
\end{aligned}
$$

where the last line is obtained through solving for G and H's convex conjugates:

$$
\begin{aligned}
G_x^*(\mu)&=\max_u\left\{\sum_{y,k}\frac{k_x\mu_{yk}}{n_x}u_{xyk}+\frac{\mu_{x0}}{n_x}u_{x0}-G_x(u)\right\}\\
H_y^*(\mu)&=\max_v\left\{\sum_{y,k}\frac{\mu_{yk}}{m_y}v_{yk}-H_y(v)\right\}.
\end{aligned}
$$

For which first order conditions are

$$
\frac{k_x\mu_{yk}}{n_x}=\frac{\exp\left(u_{xyk}\right)}{\sum_y\sum_k\exp\left(u_{xyk}\right)}\text{ and }\frac{\mu_{yk}}{m_y}=\frac{\exp\left(v_{yk}\right)}{\sum_k\exp\left(v_{yk}\right)}
$$

Which ensures that μ is feasible, i.e. satisfies margin equations 8, otherwise the value of the social planner problem is $+\infty$.

Problem (7)'s objective function is strictly concave and the maximization set defined by the margin equations (8) is compact. Therefore the maximum exists and is unique. □

Proof. Proposition 2

Let $Z_x=\max_{y\in\mathcal{Y},k\in\mathcal{K}}\left\{u_{xyk}+\varepsilon_{yk},\varepsilon_0\right\}$ and $Z_y=\max k\in\mathcal{K}v_{yk}+\eta_k$. The proof consists in showing that Z_x and Z_y follow Gumbel distributions with expectation $1+\log\sum_{y\in\mathcal{Y},k\in\mathcal{K}}\exp(u_{xyk})$

and $\log \sum_{k\in\mathcal{K}} \exp(v_{yk})$ respectively, by expressing $\mathbb{P}\left[Z_x \leq c\right]$ and $\mathbb{P}\left[Z_y \leq c\right]$.

$$\begin{aligned}
\mathbb{P}\left[Z_x \leq c\right] &= \mathbb{P}\left[\max_{y\in\mathcal{Y},k\in\mathcal{K}} \{u_{xyk} + \varepsilon_{yk}, \varepsilon_0\} \leq c\right] \\
&= \mathbb{P}\left[u_{xyk} + \varepsilon_{yk} \leq c \, \forall y, k \text{ and } \varepsilon_0 \leq c\right] \\
&= \left(\prod_{y\in\mathcal{Y},k\in\mathcal{K}} \mathbb{P}\left[u_{xyk} + \varepsilon_{yk} \leq c\right]\right) \times \mathbb{P}\left[\varepsilon_0 \leq c\right] \\
&= \left(\prod_{y\in\mathcal{Y},k\in\mathcal{K}} \mathbb{P}\left[\varepsilon_{yk} \leq c - u_{xyk}\right]\right) \times \mathbb{P}\left[\varepsilon_0 \leq c\right] \\
&= \exp\left(-\exp\left(-c\right)\right) \prod_{y\in\mathcal{Y},k\in\mathcal{K}} \exp\left(-\exp\left(u_{xyk} - c\right)\right) \\
&= \exp\left(-\exp(-c) - \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \exp\left(u_{xyk} - c\right)\right) \\
&= \exp\left(-\exp(-c)\left(1 + \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \exp\left(u_{xyk}\right)\right)\right) \\
&= \exp\left(-\exp\left(-c + \log\left(1 + \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \exp(u_{xyk})\right)\right)\right)
\end{aligned}$$

where the transition between rows two and three rests on ε being iid distributed, and row five rests on the extreme value 1 assumption. Hence up to the Euler-Mascheroni constant, Z_x follows a Gumbel distribution with expectation $\log\left(1 + \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \exp(u_{xyk})\right)$. The reasoning for Z_y follows the same steps. $\square$

B Data Appendix

I use *Quadros de Pessoal*, a matched employer-employee dataset provided by the Portuguese National Institute (Instituto Nacional de Estatística, INE). *Quadros de Pessoal* is issued yearly from 1991 to 2017[10], based on firms declarations on their characteristics and their employees'. All firms are surveyed and completing the survey is mandatory. Both workers and firms are identified across time by a unique identifier.

Worker's type are their gender and education level. Worker education is provided as a 3 digits classification, out of which I aggregate 3 levels: tertiary education (vocational 2 year degrees, bachelor, master and PhD), or secondary education (completed high school, some higher education), and no degree (no schooling, or completed schooling below high school). I

[10] There is no data in 2001

exclude from the sample any worker whose education level is unknown (3.9% of observations per year on average). Gender is split into two categories, male and female, and used as is.

The data provides average monthly earnings, that account for bonuses and extra hours earnings. Number of hours is provided as the baseline number of hours in the contract, plus any extra hours worked (averaged overt the year). The sum of base and extra hours is used as the measure for number of hours worked per month. I divide monthly earnings by monthly hours to obtain a measure of hourly wage, and deflate it. I exclude from the sample any worker who has earned zero wage over the year (11.1% of the sample). These are mainly workers on sick leave, maternity leave, or sabbatical that do not contribute to firm production in that year. Workers below the hourly minimum wage are also excluded (4.9%). Observations with unknown education level, invalid or duplicated work ID are dropped (1.1%, 1.2% and .4% of the sample respectively). Part-time workers are dropped (12.7% of the sample), as well as self-employed (if the firm has only one employee), The rationale behind not considering self-employed is that many of self-employed workers actually work as consultants for a firm, with no way to link them. Self-employed workers amount for 2.4% of the sample. I also exclude from the sample any workers who are strictly under 16 or above 68 (the retirement age in Portugal).

Unemployment rates by gender and education level are available in the World Bank's database. The data provides unemployment rates by advanced, intermediate and basic education, which correspond to the previously defined levels. The numbers of unemployed workers are computed from the unemployment rates and the number of employed workers from each type.

C Estimation Appendix

C.1 Firm Classification

The *kmeans*' group assignment works in two steps; first choose a number of clusters (or types) Y; and second, assign each individual firm j to a Type y, based on their wage deciles $w_j = [w_{j,10}, \ldots, w_{j,90}]$. The *kmeans* algorithm defines types y by defining group clusters $C_1, \ldots, C_Y$ that solve the following optimization problem:

$$\min_{C_1,\ldots,C_Y} \sum_{y=1}^{Y} \sum_{j \in C_y} \|w_j - \omega_y\|^2$$

where $\omega_y = \frac{1}{|C_y|} \sum_{j \in C_y} w_j$ is group C_y's centroid.

The *kmeans* optimization consists of finding the groups $C_1, \ldots, C_Y$ and allocation of firms

j to these groups such that the sum of the squared distances between each firm's wage deciles and the group's centroid ω_y is minimized. The algortihm outputs both a firm classification into types y and the centroids ω_y of each type. I test a different number of types Y in 2017 and use the elbow method to choose the best number $Y = 7$. I then use the types and their centroids computed in 2017 to classify firms in previous years. In the years prior to 2017, I assign a firm to a group C_y if its centroid is the closest, in Euclidian terms, to the firm's normalized wage deciles. Since the types are defined up to labelling, I rank them by firm average wage to label them from 1 to 7 (Group 1 pays the lowest average wage, and Group 7 the highest). In the literature, firms belonging to groups with high average wage (Groups 6 and 7) are referred to as high type, and firms belonging to groups with low average wage (Groups 1,2,3) are low type. The middle Groups 4 and 5 are referred as medium.

C.1.1 Matching Moments

Under the parametrization described in sub-section 3.2, Proposition 3 provides an expression for the predicted matching μ as a function of U, V and ϕ:

$$\begin{aligned}\mu_{yk}(\phi, U, V) = \exp\Bigg(\frac{1}{1+\sum_x k_x}\Bigg(&\sum_{x\in\mathcal{X}_k}\phi^1_{xy}k_x + \sum_{x,x'\in\mathcal{X}_k}\phi^2_{xx'y}k_x\log k_{x'}\\ &-\sum_{x\in\mathcal{X}_k}k_xU_x - V_y - \sum_{x\in\mathcal{X}_k}k_x\log k_x\Bigg)\Bigg) \quad \forall y\in Y, k\in\mathcal{K}\\ \mu_{x0}(U) = \exp\left(-U_x\right) \quad \forall x\in\mathcal{X}&\end{aligned} \tag{13}$$

where $\mathcal{X}_k$ are the set of types $x\in\mathcal{X}$ such that $k_x > 0$. The parametrization suggests natural matching patterns to match:

$$\begin{aligned}&\sum_{k\in\mathcal{K}} k_x\mu_{yk}(\phi, U, V) \quad \forall x\in\mathcal{X}, y\in\mathcal{Y}\\ &\sum_{\substack{k\in\mathcal{K}\\ k_{x'}>0}} k_x\log k_{x'}\mu_{yk}(\phi, U, V) \quad \forall x, x'\in\mathcal{X}, y\in\mathcal{Y}\end{aligned} \tag{14}$$

Let $\tilde{\mu} = (\tilde{\mu}_{yk})_{y,k}$ be the observed matching distribution; that is each $\tilde{\mu}_{yk}$ records the number of Type y firms employing workforce k in the data. Then, the observed counterparts to the matching moments (14) are simply

$$\sum_{k\in\mathcal{K}} k_x \tilde{\mu}_{yk} \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

$$\sum_{\substack{k\in\mathcal{K}\\ k_{x'}>0}} k_x \log k_{x'} \tilde{\mu}_{yk} \quad \forall x, x' \in \mathcal{X}, y \in \mathcal{Y}$$

Similar to $\tilde{\mu}$, let $\tilde{\mu}_0 = (\tilde{\mu}_{x0})_x$ be the count of observed unemployed workers of Type x. Additional predicted moments are given by the margin equations:

$$\sum_{y\in\mathcal{Y},k\in\mathcal{K}} k_x \mu_{yk}(\phi, U, V) + \mu_{x0}(U) = n_x \text{ and } \sum_{y\in\mathcal{Y}} \mu_{yk}(\phi, U, V) = m_y$$

for whose the observed counterparts are $\sum_{y\in\mathcal{Y},k\in\mathcal{K}} k_x \tilde{\mu}_{yk} + \tilde{\mu}_{x0} = \tilde{n}_x$ and $\sum_{y\in\mathcal{Y}} \tilde{\mu}_{yk} = \tilde{m}_y$

Under Assumption 1, fitting the predicted and observed matching moments to estimate (ϕ, U, V) is equivalent to running a Poisson regression that fits the functional form (13) on the observed matching $\tilde{\mu}$ and unemployed $\tilde{\mu}_0$, as stated by Theorem 2. It generalizes Galichon and Salanié (2024)'s, who provide a proof in the one-to-one case. The Theorem states that the moment-matching estimator described in subsection C.1.1 is the solution to an optimization problem that amounts to the Poisson regression of 1. observed matching $\tilde{\mu}_{yk}$ on the surplus basis functions $\left(\sum_{y'} \mathbb{1}_{y'=y} k_x\right)_{x,y}$ and $\left(\sum_{y'} \mathbb{1}_{y'=y} k_x \log k_{x'}\right)_{x,x',y}$, $(-k_x)_x$ and $\left(-\sum_{y'} \mathbb{1}_{y'=y}\right)_y$, with offset $\sum_x k_x \log k_x$ and weight $(1 + \sum_x k_x)$ and 2. observed unemployed $\tilde{\mu}_{x0}$ on fixed effects $(-\sum_{x'} \mathbb{1}_{x'=y})_x$ with offset 0 and weight 1.

Theorem 2. *Under Assumption 1, the moment-matching estimator $\left(\hat{\phi}, \hat{U}, \hat{V}\right)$ is the solution to*

$$\begin{aligned}\max_{\phi,U,V} \sum_{y\in\mathcal{Y},k\in\mathcal{K}} \tilde{\mu}_{yk} \Bigg(\sum_{x\in\mathcal{X}_k} \phi^1_{xy} k_x + \sum_{x,x'\in\mathcal{X}_k} \phi^2_{xx'y} k_x \log k_{x'} - \sum_{x\in\mathcal{X}_k} k_x U_x - V_y - \sum_{x\in\mathcal{X}_k} k_x \log k_x \Bigg) \\ + \sum_{x\in\mathcal{X}} \tilde{\mu}_{x0}(-U_x) \\ - \sum_{y\in\mathcal{Y},k\in\mathcal{K}} (1 + \sum_{x\in\mathcal{X}} k_x) \mu_{yk}(\phi, U, V) - \sum_{x\in\mathcal{X}} \mu_{x0}(U)\end{aligned}$$

where $\tilde{\mu}$ is the observed matching and $\mu_{yk}(\phi, U, V)$ is defined as in (13).

Proof. The equalities between the matching moments to their data counterparts:

$$\begin{aligned}
\sum_{k\in\mathcal{K}} k_x \mu_{yk}(\phi, U, V) &= \sum_{k\in\mathcal{K}} k_x \tilde{\mu}_{yk} \quad \forall x \in \mathcal{X}, y \in \mathcal{Y} \\
\sum_{k\in\mathcal{K}} k_x \log k_{x'} \mu_{yk}(\phi, U, V) &= \sum_{k\in\mathcal{K}} k_x \log k_{x'} \tilde{\mu}_{yk} \quad \forall x, x' \in \mathcal{X}, y \in \mathcal{Y} \\
\sum_{y\in\mathcal{Y}, k\in\mathcal{K}} k_x \mu_{yk}(\phi, U, V) + \mu_{x0}(U) &= \sum_{y\in\mathcal{Y}, k\in\mathcal{K}} k_x \tilde{\mu}_{yk} + \tilde{\mu}_{x0} \quad \forall x \in \mathcal{X} \\
\sum_{k\in\mathcal{K}} \mu_{yk}(\phi, U, V) &= \sum_{k\in\mathcal{K}} k_x \tilde{\mu}_{yk} \quad \forall y \in \mathcal{Y}
\end{aligned}$$

are the First Order Conditions of the optimization problem in Theorem 2. □

Running the Poisson regression provides estimates $\left(\hat{\phi}, \hat{U}, \hat{V}\right)$. Let $\hat{\mu}_{yk} = \mu_{yk}(\hat{\phi}, \hat{U}, \hat{V})$ and $\hat{\mu}_{x0} = \mu_{x0}(\hat{U})$ be the predicted matching and unemployed distributions. The variance-covariance matrix is

$$V^{\phi,U,V} = A^{-1} B A^{-1}$$

where $A = Z^\top W \operatorname{diag}\left(\mu(\Phi, U, V)\right) WZ$, $B = Z^\top W \operatorname{diag}(\tilde{\mu} - \tilde{\mu}\tilde{\mu}^\top) WZ$, W is a diagonal matrix with entries equal to the number of individual in the match between firm y and workforce k: $1 + \sum_{x\in\mathcal{X}} k_x$ and Z is the matrix of regressors from (13).

C.1.2 Wage Moments

Under the model's parametrization, Proposition 3 also provides a functional form for wage w:

$$w_{xyk}(\alpha) = \hat{U}_x - \alpha^1_{xy} - \sum_{x\in\mathcal{X}_k} \alpha^2_{xx'y} \log k_x + \log k_x + \log \hat{\mu}_{yk}$$

where $\left(\hat{\phi}, \hat{U}, \hat{V}\right)$ were estimated in the first step. Define the following wage moments for estimating α:

$$\begin{aligned}
&\sum_{k\in\mathcal{K}} k_x \hat{\mu}_{yk} w_{xyk}(\alpha) \quad \forall x \in \mathcal{X}, y \in \mathcal{Y} \\
&\sum_{\substack{k\in\mathcal{K} \\ k_{x'}>0}} k_x \log k_{x'} \hat{\mu}_{yk} w_{xyk}(\alpha) \quad \forall x, x' \in \mathcal{X}, y \in \mathcal{Y}
\end{aligned} \tag{15}$$

The data provide observed wage $\tilde{w}_{ijk(j)}$, where i is a worker of Type x, and j a firm of Type y employing workforce $k(j)$. The observed counterpart to the first wage moment is simply the sum of all wages of Type x workers employed in Type y firms, and the counterpart to the second moment is the sum of all wages of Type x workers employed in Type y firms, weighted by the log of the number of workers of Type x' in the workforce:

$$\sum_{\substack{i:x_i=x \\ j:y_j=y}} k(j)_x \tilde{w}_{ijk(j)} \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$
$$\sum_{\substack{i:x_i=x \\ j:y_j=y}} k(j)_x \log k(j)_{x'} \tilde{w}_{ijk(j)} \quad \forall x, x' \in \mathcal{X}, y \in \mathcal{Y}$$

Predicted wages are fitted to their observed counterpart using the two-steps GMM method. Fitting the wage moments yields estimated parameters $\hat{\alpha}$. Their variance-covariance matrix is

$$V^{\alpha} = H^{-1} G_{\alpha}^{\top} \Omega M \Omega G_{\alpha} H^{-1}$$

and the covariance with (ϕ, U, V) is

$$V^{\alpha,\phi,U,V} = -\frac{n}{N} V^{\phi,U,V} G_{\Phi,U,V}^{\top} \Omega G_{\alpha} H^{-1}$$

where n is the number of wage observations, N is the number of matching observations, $H = G_{\alpha}^{\top} \Omega G_{\alpha}$,$G_{\alpha}$ is the Jacobian of the wage moments with respect to α, $G_{\Phi,U,V}$ is the Jacobian of the wage moments with respect to (ϕ, U, V), and Ω is the weight matrix used in the second step of the GMM estimation, that is

$$\Omega^{-1} = \tilde{V} + \frac{n}{N} G_{\phi,U,V} V^{\phi,U,V} G_{\phi,U,V}^{\top}$$

where $\tilde{V}$ is the variance covariance matrix of the wage moments at the intermediate estimates for α obtaine din the GMM's first step.

C.1.3 Workforce Grid Dimension Reduction

The space of workforce grids $\mathcal{K}$ is the Cartesian product of the sets $\{0, \ldots, \tilde{n}_x\}$ for all $x \in \mathcal{X}$, where $\tilde{n}_x$ is the observed number of workers of Type x in the data. In practice, this is a very high-dimensional object, which makes computing the predicted matching and wage moments (14) and (15) computationally challenging. To alleviate computer memory issues and speed up estimation, I instead use a subset $\mathcal{K}^c$ of the entire workforce grid. To build subset $\mathcal{K}^c$, I leverage observation of the grid $\tilde{\mathcal{K}}$ from the data, which is simply the set of all workforces k that are observed in the matched employer-employee dataset over all available years, from 1991 to 2017. In total there are 584,953 different workforce combinations that form over this period. This is already a subset of $\mathcal{K}$, as not all workforce combinations are realized. However, using $\tilde{\mathcal{K}}$ would present two problems; first, it is still too large a grid for an estimation in reasonable time; and second, it would result in few zero observations for matching $\tilde{\mu}$, given that it was selected to be observed workforce combinations. Selecting on non-zero observations in a Poisson regression is known to bias the estimates (Grogger and Carson (1991)). Instead, I draw a sample $\mathcal{K}^c$ of size 50,000 from $\tilde{\mathcal{K}}$ and run a hierarchical clustering algorithm (Murtagh and Contreras (2011)) to assign all grid points in $\tilde{\mathcal{K}}$ to grid points in $\mathcal{K}^c$. Doing so obtains observed matching $\tilde{\mu}_{yk}$ on the reduced grid $\mathcal{K}^c$, which I then use in the Poisson regression in the first estimation step. In the second estimation step, the predicted wage moments (15) are computed on the reduced grid $\mathcal{K}^c$ as well.

I check whether using $\mathcal{K}^c$ is not too restrictive by comparing the between worker type, firm type and workforce variance of wages over the full grid $\tilde{K}$ and the reduced grid $\mathcal{K}^c$. Wage sample variance can be decomposed as follows:

$$V_{\tilde{\mathcal{K}}}\left[\tilde{w}\right] = \underbrace{\frac{1}{\tilde{N}-1} \sum_{k \in \tilde{\mathcal{K}}, y \in \mathcal{Y}} \sum_{x \in \mathcal{X}_k} k_x \tilde{\mu}_{yk} \left(\overline{\tilde{w}}_{xyk} - \overline{\tilde{w}}\right)^2}_{\text{between}} + \underbrace{\frac{1}{\tilde{N}-1} \sum_{k \in \tilde{\mathcal{K}}, y \in \mathcal{Y}} k_x \tilde{\mu}_{yk} \sum_{x \in \mathcal{X}_k} \sigma^2_{xyk}}_{\text{within}}$$

where $\overline{\tilde{w}}$ is the average wage over the entire sample, and $\overline{\tilde{w}}_{xyk}$ and σ^2_{xyk} are the sample average and variance of wages of Type x workers employed in Type y firms within workforce k. $\tilde{N}$ is the total number of wage observations in the sample. The between variance captures the variation in wages across worker types, firm types and workforces, while the within variance captures the variation in wages within each group. The model predicts a distribution of wages $(w_{xyk})_{x,y,k}$ that can rationalize the between variance observed in the data, while the within variance is treated as a residual. A criterion for whether the approximation of $\mathcal{K}$ with $\mathcal{K}^c$ is a good one is whether it captures the original between variance over $\tilde{K}$ correctly, i.e. whether the between variance computed over $\mathcal{K}^c$ is close to the one computed over $\tilde{K}$. I compute both in 1992 and 2017. In 1992, the between variance over $\tilde{K}$ explains 56.7% of the total variance, while the between variance over $\mathcal{K}^c$ explains 56.1%. In 2017, these shares are 58.6% and 57.6%, respectively. I conclude that I do not lose much variation in wages by using the reduced workforce grid $\mathcal{K}^c$.

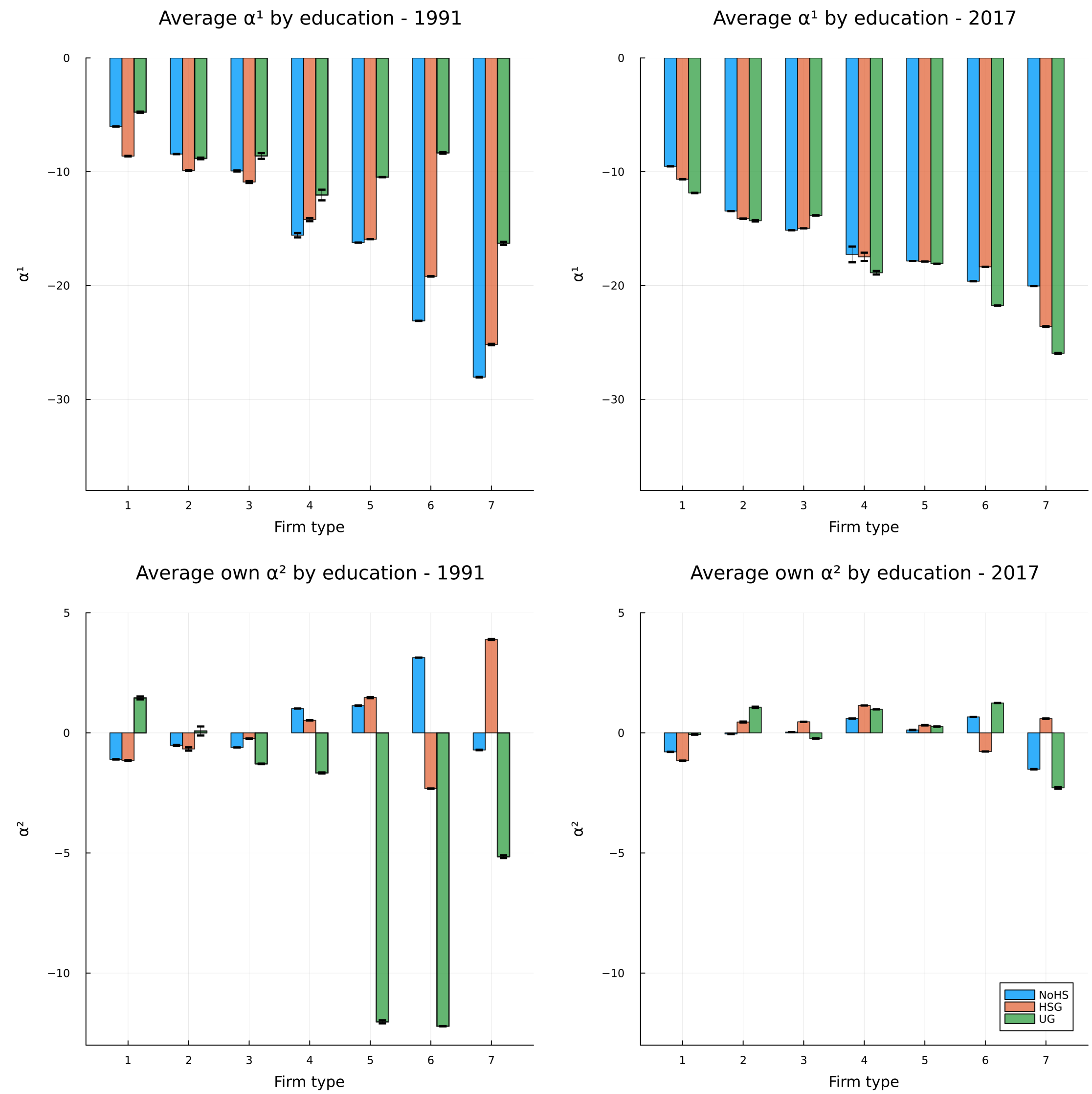
D Results Appendix

D.1 Estimated Parameters

D.1.1 Own Parameters

Figure 6 is the analogue to Figure 8 for amenities, i.e. it plots the average own contribution to amenities by education level. There are four observations to be made from this figure. First, perceived first-order amenities are significantly negative for all education levels across firm types. These are to be compared to worker's outside option, which is to remain unemployed, meaning there is a significant disutility to work. All other things equal, this results in higher wages. Second, perceived amenities are worse in high-type than low-type firms. Third, first-order perceived amenities decrease between 1991 and 2017 for university graduates, while they increase for high school graduates and workers with no degree. All other things equal this pressures the education wage premia upwards over time. Finally, second-order perceived amenities are very low for university graduates in high-type firms in 1991. Because they are most productive in these firms (as shown in Figure 8), this puts an upward pressure on their wage compared to other education levels in 1991.

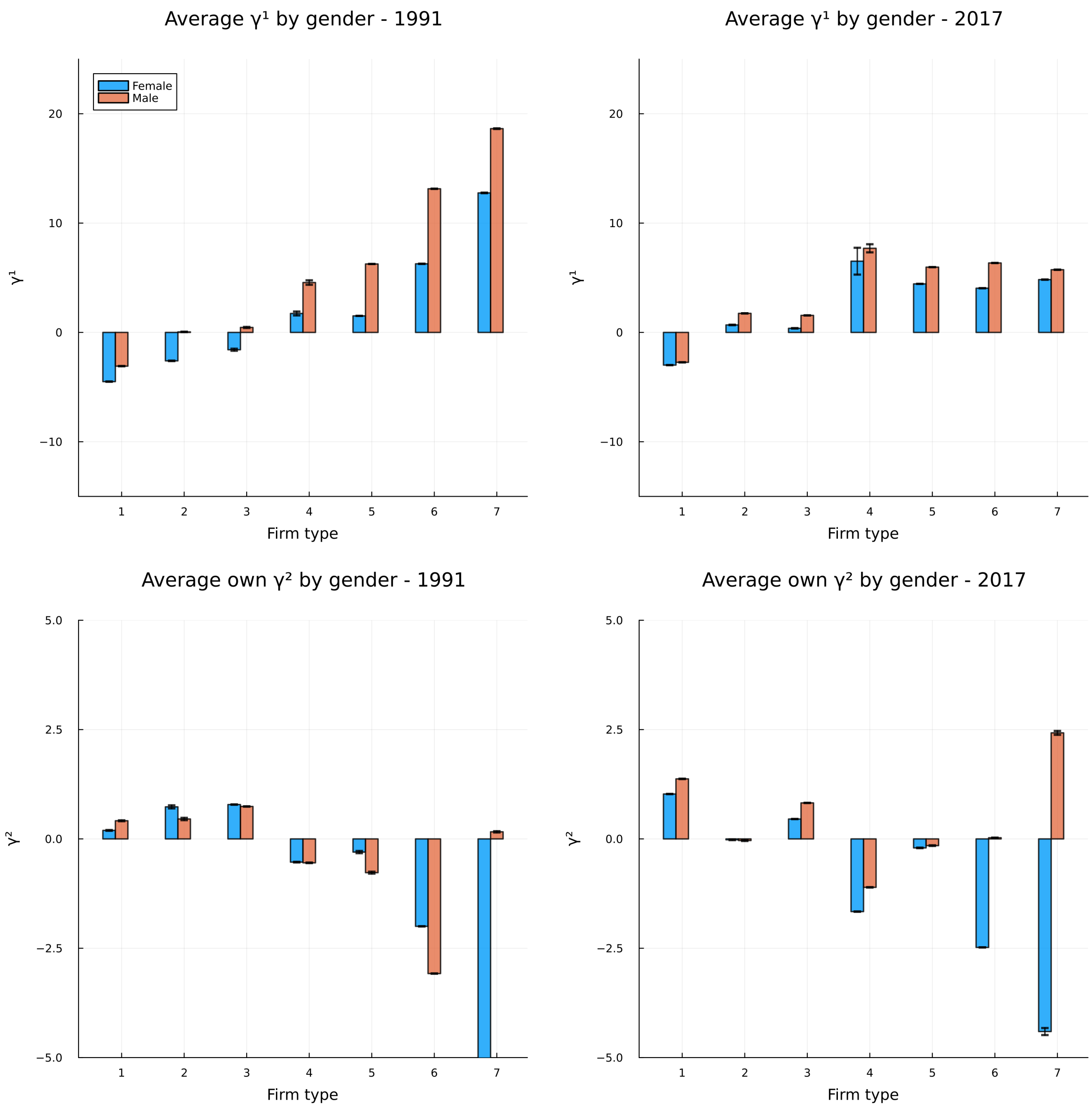
Figure 6: Own Contributions to Amenities - Education



Source: *Quadros de Pessoal*. Author's own calculations from estimated parameters. Weighted averages by education are computed over both gender, e.g. for a given firm cluster y and education level NoHS: $\alpha^1_{\text{NoHS},y} = \frac{m_{\text{F-NoHS}}\alpha^1_{\text{F-NoHS},y} + m_{\text{M-NoHS}}\alpha^1_{\text{M-NoHS},y}}{m_{\text{F-NoHS}} + m_{\text{M-NoHS}}}$ and $\alpha^2_{\text{NoHS},textNoHS,y} = \frac{m_{\text{F-NoHS}}\alpha^1_{\text{F-NoHS},textNoHS,y} + m_{\text{M-NoHS}}\alpha^1_{\text{M-NoHS,M-NoHS},y}}{m_{\text{F-NoHS}} + m_{\text{M-NoHS}}}$. Euros per hours. Confidence intervals are provided at the 95% level.

Figure 7 is the analogue to Figure 7 with averages of own contribution to productivity being taken over gender instead of education. The first-order contributions of both men and women are increasing in firm type in both 1991 and 2017. In 1991, men's first-order contributions are higher than women's in all firm types. In 2017 it is still the case but the gap is much smaller. This at least partly reflects a composition effect over education levels which evolves over time as women's education progresses. The most important difference between men and women is in the second-order own contributions: women tend to be strong substitute with each other in high type firms, both in 1991 and 2017. A number of explanations can be put forward for this: it could be due to gender selection into occupations, which due to a limited number of female-oriented occupations limit the number of women employed i high-type firms, or it could be discrimination on the firm side if it does not want to hire to many women.

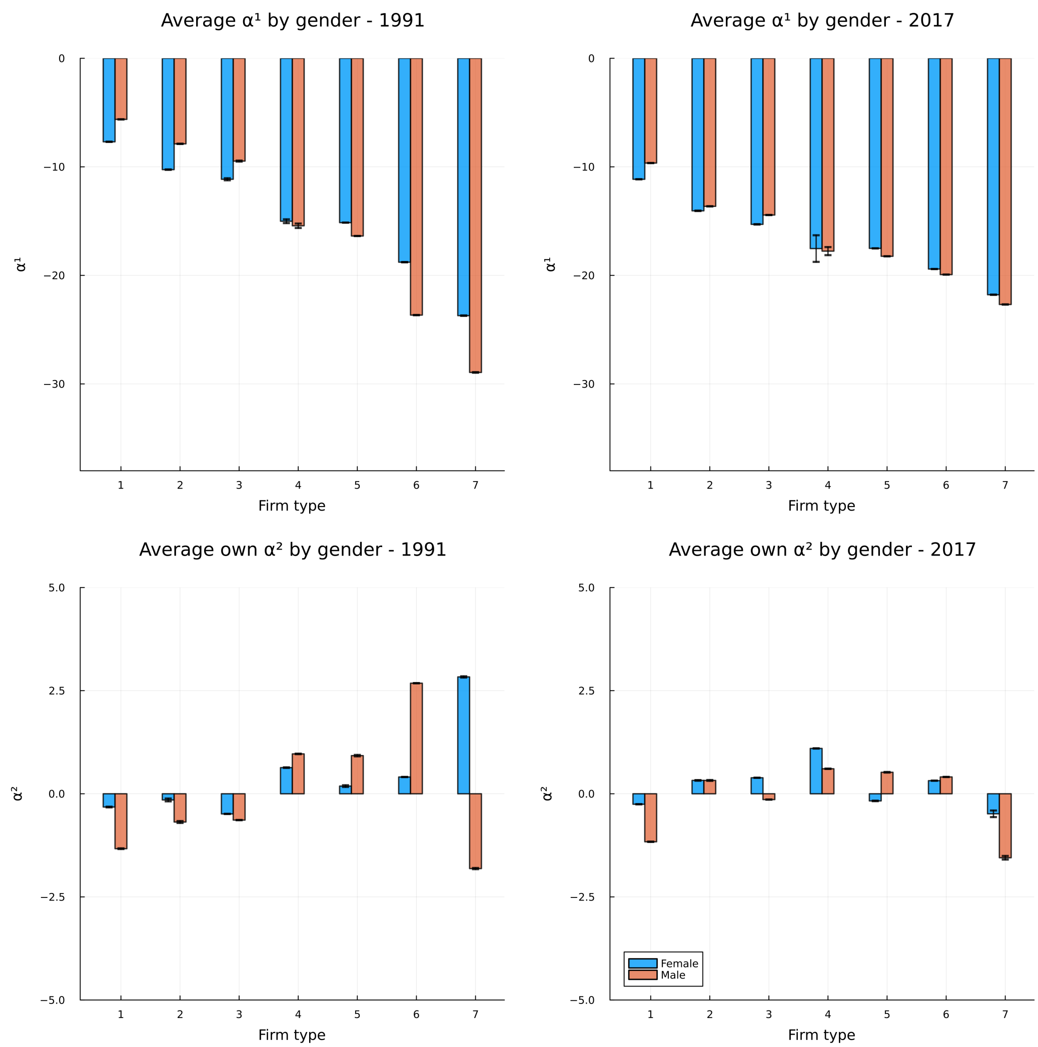
Figure 7: Own Contributions to Productivity - Gender



Source: *Quadros de Pessoal*. Author's own calculations from estimated parameters. Weighted averages by gender are computed over all education levels, e.g. for a given firm cluster y and gender F: $\gamma^1_{\text{F},y} = \frac{m_{\text{F-NoHS}}\gamma^1_{\text{F-NoHS},y}+m_{\text{F-HSG}}\gamma^1_{\text{F-HSG},y}+m_{\text{F-UG}}\gamma^1_{\text{F-UG},y}}{m_{\text{F-NoHS}}+m_{\text{F-HSG}}+m_{\text{F-UG}}}$. Euros per hours. Confidence intervals are provided at the 95% level.

Figure 8 presents average own contributions to amenities by gender. The first-order contributions are negative for both genders, more so for men than women, and for high-type firms than low-type firms. They all increase between 1991 and 2017, which puts a downward pressure on both men and women's wages. The second-order contributions are overall positive for women, which decreases their wages in firms where they are well-represented. They are bell-shaped over firm types for men, who all other things equal will therefore be paid more in type 7 firms (where they are also most productive), than other firm types.

Figure 8: Own Contributions to Amenities - Gender

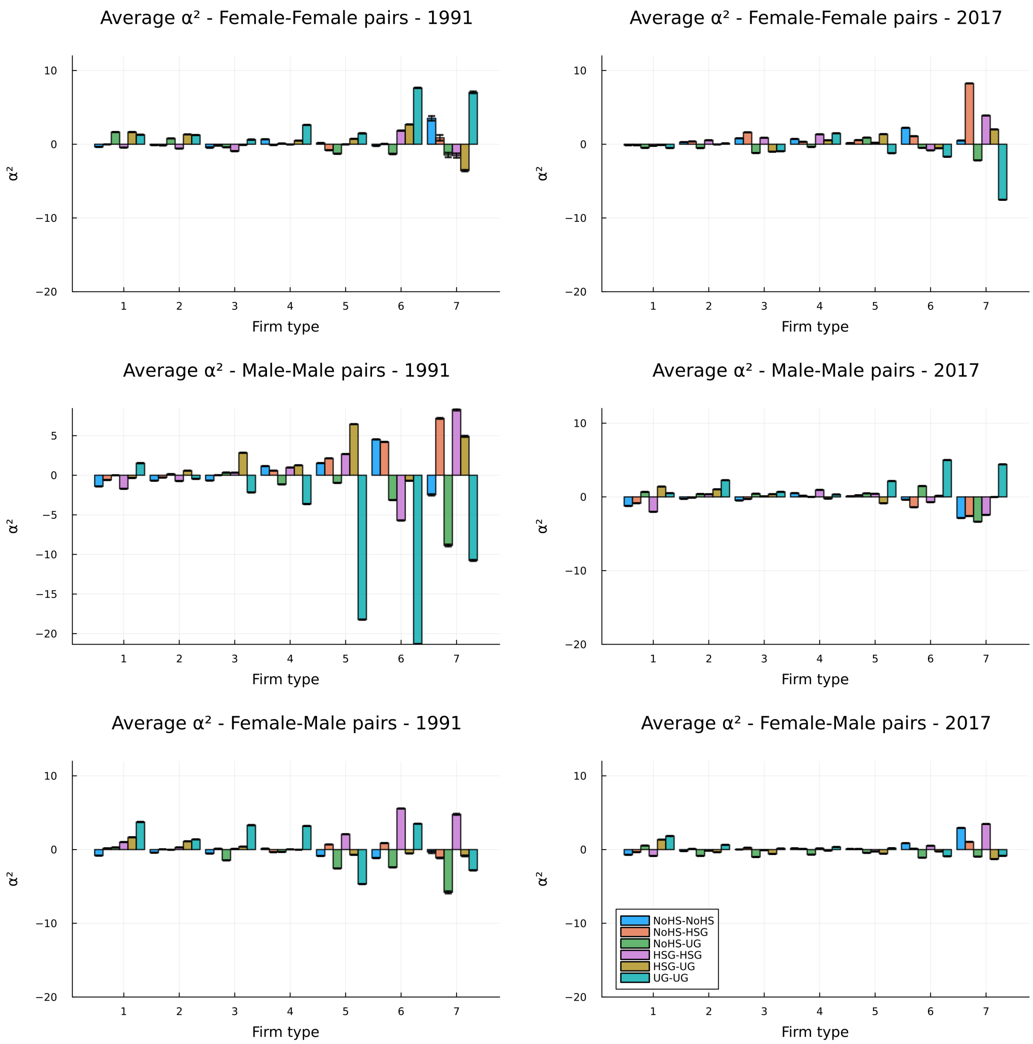


Source: *Quadros de Pessoal*. Author's own calculations from estimated parameters. Weighted averages by gender are computed over all education levels, e.g. for a given firm cluster y and gender F: $\alpha^1_{\mathrm{F},y} = \frac{m_{\text{F-NoHS}}\alpha^1_{\text{F-NoHS},y}+m_{\text{F-HSG}}\alpha^1_{\text{F-HSG},y}+m_{\text{F-UG}}\alpha^1_{\text{F-UG},y}}{m_{\text{F-NoHS}}+m_{\text{F-HSG}}+m_{\text{F-UG}}}$. Euros per hours. Confidence intervals are provided at the 95% level.

D.1.2 Cross Parameters

Figures 9 and 10 plot the estimated second degree parameters of amenities and productivity, respectively. To reduce the number of parameters to plot, these are averaged out over a given education and gender pair. The left panel shows these averages in 1991, the right panel in 2017. There are three education pairs: Female-Female (first row), Male-Male (second row) and Female-Male (third row). Each bar color represents one of the six education pair: NoHS-NoHS, NoHS-HSG, NoHS-UG, HSG-HSG, HSG-UG, UG-UG. In Figure 9, the second degree parameters of amenities are overall more dispersed in 1991 than in 2017, especially among male and university graduates pairs in high firm types. Male university graduates in particular strongly dislike working with each other in firm types 5 to 7, while women university graduates prefer it. In 2017, the tastes are reversed: male university graduates are happy to work together while female university graduates are not. Male-Female pairs are much closer to zero both in 1991 and 2017, signalling the two genders are indifferent to working with one another regardless of their education level.

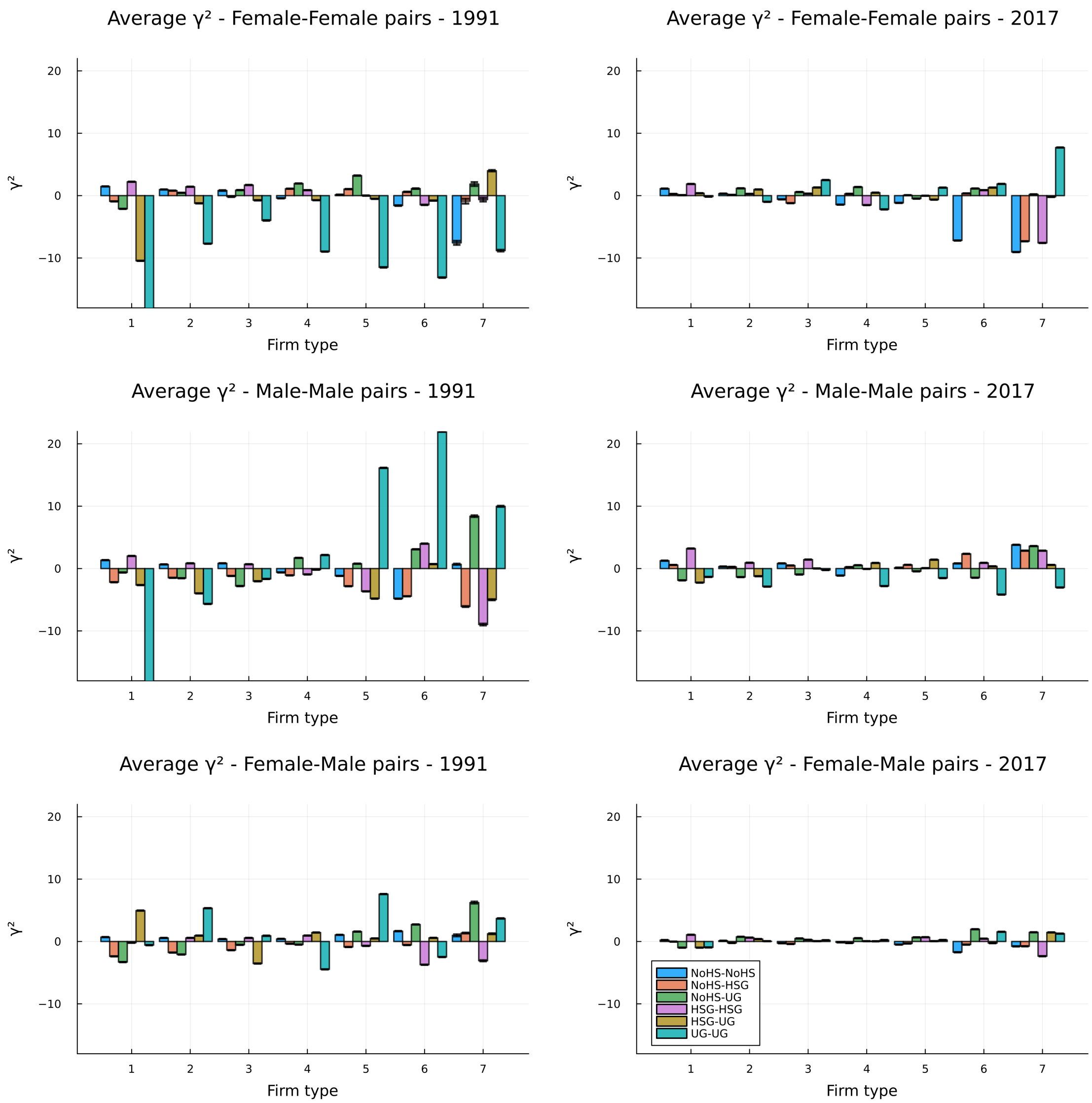
Figure 9: Cross-Contributions to Amenities



Source: *Quadros de Pessoal*. Author's own calculations from estimated parameters. Weighted averages by gender and education pairs are computed as follows: for a given firm cluster y, gender pair F-M and education pair NoHS-HSG: $\alpha^2_{\text{F-M,NoHS-HSG},y} = \frac{m_{\text{F-NoHS}} \times m_{\text{M-HSG}}\left(\alpha^2_{\text{F-NoHS,M-HSG},y} + \alpha^2_{\text{M-HSG,F-NoHS},y}\right) + m_{\text{F-HSG}} \times m_{\text{M-NoHS}}\left(\alpha^2_{\text{F-HSG,M-NoHS},y} + \alpha^2_{\text{M-NoHS,F-HSG},y}\right)}{2m_{\text{F-NoHS}} \times m_{\text{M-HSG}} + 2m_{\text{F-HSG}} \times m_{\text{M-NoHS}}}$, where m are the masses of each type. Confidence intervals are provided at the 95% level.

In Figure 10, the more dispersed second degree parameters of productivity in 1991 are again male and female university graduates pair: men university graduates working together have a strong positive contribution to high type firms' productivity, while female university graduates have a negative contribution. The effects wear off in 2017, when interdependency patterns in productivity are much smaller for most gender and education pairs. The only exception are women with no degree or a high school degree in high firm types which contribute negatively to productivity.

Figure 10: Cross-Contributions to Productivity



Source: *Quadros de Pessoal.* Author's own calculations from estimated parameters.Weighted averages by gender and education pairs are computed as follows: for a given firm cluster y, gender pair F-M and education pair NoHS-HSG: $\gamma^2_{\text{F-M,NoHS-HSG},y} = \frac{m_{\text{F-NoHS}} \times m_{\text{M-HSG}}\left(\gamma^2_{\text{F-NoHS,M-HSG},y} + \gamma^2_{\text{M-HSG,F-NoHS},y}\right) + m_{\text{F-HSG}} \times m_{\text{M-NoHS}}\left(\gamma^2_{\text{F-HSG,M-NoHS},y} + \gamma^2_{\text{M-NoHS,F-HSG},y}\right)}{2m_{\text{F-NoHS}} \times m_{\text{M-HSG}} + 2m_{\text{F-HSG}} \times m_{\text{M-NoHS}}}$, where m are the masses of each type. Confidence intervals are provided at the 95% level.

D.2 Counterfactuals

D.2.1 Averaged out workers' preferences

The fourth counterfactual demonstrates the importance of estimated workers' non-wage preferences (or perceived amenities) on sorting and the education wage premium. It sets all preference parameters α^1 and α^2 to their average estimated value over firm types. Workers of different education levels and gender still perceive different amenity levels, but these are now the same across firm types. These levels vary over time as estimated. Subsection 4.3 establishes that all worker types dislike being employed in high-type firms more than low-type firms. In 1991, this is especially true of men and workers who have no degree. In 2017 the gaps are less pronounced. There are two competing channels whereby workers' non-wage preferences can affect the education wage premia. First, workers' strong dislike for high-type firms versus low and medium type firms should reduce sorting into high-type firms. The fact that these firms are most productive for educated workers, especially in 1991, reduces the education wage premia. Second, for the workers who are sorted into high-type firms, their low perceived amenities drives their wages up. Since educated workers in 2017 have a stronger distaste for high-type firms than less-educated workers, this should drive the education wage premia up. Table 11 shows the predicted and counterfactual sorting into high-type firms. Averaging out workers' preferences erases their strong relative dislike for high-type firms. As a result, educated workers sort much more into high-type firms in the counterfactual than in the predicted baseline: in 1991, 98.6% of the university graduates and 65.5% of the high school graduates were employed in high-type firms, which are also the most productive for educated workers. In 2017, the relative productivity of low-type firm rose, which is why the sorting in high-type firms is less pronounced, but still quite high (70.0% instead of 36.6% for university graduates, and 77.2% instead of 17.1% for high school graduates).

Table 11: Averaged worker preferences - sorting

	% UG in high firm types		% HSG in high firm types	
	1991	2017	1991	2017
Predicted	48.2	36.6	37.6	17.1
Counterfactual	98.6	70.0	65.5	77.2

Source: *Quadros de Pessoal.* Author's calculations from model estimates. High types: Types 6-7. UG: university graduates. HSG: high school graduates. Shares are computed among all employed workers of a given education level.

From Table 12, both university and high school wage premia are reduced by equalizing workers' preferences across firms in 1991. The reason is that less-educated workers originally have a strong aversion to high-type firms; once preferences are equalized, their relative dislike shifts toward other firm types, raising their wages and compressing the gaps. In 2017, the counterfactual high school wage premium is close to the predicted baseline, but the university wage premium is much lower (.14 instead of .79). This confirms that the university graduates' strong dislike for high-type firms in 2017 weakens their sorting in these firms (as shown in Table 11), but also places upwards pressure on their wages.

Table 12: Averaged worker preferences - wage gaps

	UG-NoHS wage gap			HSG-NoHS wage gap		
	1991	2017	Δ	1991	2017	Δ
Predicted	1.1	.79	-.33	.28	.26	-.02
Counterfactual	.69	.14	-.55	-.44	.29	.72

Source: *Quadros de Pessoal*. Author's calculations from model estimates. UG: university graduates. HSG: high school graduates. NoHS: less-educated workers. Wage gaps are computed as the difference of the log of the average wage of a given education level and the log of the average wage of less-educated workers (without a high school degree, NoHS). The average wage is computed over all employed workers of a given education level.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Azevedo, E. M. and J. W. Hatfield (2018): "Existence of Equilibrium in Large Matching Markets With Complementarities," .

Babet, D., O. Godechot, and M. G. Palladino (2022): "In the Land of AKM: Explaining the Dynamics of Wage Inequality in France," *Working Paper*.

Beaudry, P., D. A. Green, and B. M. Sand (2015): "The Great Reversal in the Demand for Skill and Cognitive Tasks," *Journal of Labor Economics*, 34, S199–S247.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

Bikhchandani, S. and J. M. Ostroy (2002): "The Package Assignment Model," *Journal of Economic Theory*, 107, 377–406.

Blair, P. Q. and D. J. Deming (2020): "Structural Increases in Demand for Skill after the Great Recession," *AEA Papers and Proceedings*, 110, 362–365.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87, 699–739.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.

Card, D. and T. Lemieux (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, 116, 705–746.

Cardoso, A. R. (2004): "Jobs for Young University Graduates: Is It Worth Having a Degree?" *IZA Discussion Paper Series*.

Centeno, M. and A. A. Novo (2014): "When Supply Meets Demand: Wage Inequality in Portugal," *IZA Journal of European Labor Studies*, 3, 23.

Che, Y.-K., J. Kim, and F. Kojima (2019): "Stable Matching in Large Economies," *Econometrica*, 87, 65–110.

Choné, P. and F. Kramarz (2021): "Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling," *CREST Working Paper*.

Choo, E. and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114, 175–201.

Fox, J. T. (2010): "Identification in Matching Games," *Quantitative Economics*, 1, 203–254.

Galichon, A. and B. Salanié (2022): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *The Review of Economic Studies*, 89, 2600–2629.

——— (2024): "Estimating Separable Matching Models," *Journal of Applied Econometrics*, 39, 1021–1044.

Goldin, C. and L. Katz (2008): *The Race Between Education and Technology.*, Belknap Press for Harvard University Press.

Gretsky, N., J. Ostroy, and W. Zame (1992): "The Nonatomic Assignment Model," *Economic Theory*, 2, 103–27.

Grogger, J. T. and R. T. Carson (1991): "Models for Truncated Counts," *Journal of Applied Econometrics*, 6, 225–238.

Instituto Nacional de Estatística (INE) (1991-2017): "Quadros de Pessoal," https://www.ine.pt, unpublished microdata accessed via application to INE, Portugal.

Katz, L. F. and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107, 35–78.

Kelso, A. S. and V. P. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50, 1483–1504.

Krueger, D., F. Perri, L. Pistaferri, and G. L. Violante (2010): "Cross-Sectional Facts for Macroeconomists," *Review of Economic Dynamics*, 13, 1–14.

Lamadon, T., J. Lise, C. Meghir, and J.-M. Robin (2024): "Labor Market Matching, Wages, and Amenities," .

Lamadon, T., M. Mogstad, and B. Setzler (2022): "Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market," *American Economic Review*.

Lindenlaub, I. (2017): "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 84, 718–789.

Lindenlaub, I. and F. Postel-Vinay (2021): "The Worker-Job Surplus," *NBER Working Paper*.

Lise, J., C. Meghir, and J.-M. Robin (2016): "Matching, Sorting and Wages," *Review of Economic Dynamics*, 19, 63–87.

Lise, J. and F. Postel-Vinay (2020): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 110, 2328–2376.

Murtagh, F. and P. Contreras (2011): "Methods of Hierarchical Clustering," .

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019): "Firming Up Inequality," *The Quarterly Journal of Economics*, 134, 1–50.

Sorkin, I. (2018): "Ranking Firms Using Revealed Preference*," *The Quarterly Journal of Economics*, 133, 1331–1393.

Verdugo, G. (2014): "The Great Compression of the French Wage Structure, 1969–2008," *Labour Economics*, 28, 131–144.

Vohra, R. V. (2011): *Mechanism Design: A Linear Programming Approach*, Econometric Society Monographs, Cambridge: Cambridge University Press.