Education Expansion, Sorting, and the Decreasing Education Wage Premium

Pauline Corblet*

November 22, 2021

JOB MARKET PAPER

Click here for the latest version

https://paulinecorblet.github.io/pdf/JMP.pdf

Abstract

This paper studies the interplay between worker supply and firm demand, and their effect on sorting and wages in the labor market. I build a model of one-to-many matching with multidimensional types in which several workers are employed by a single firm. Matching is dictated by worker preferences, their relative productivity in the firm, and substitution patterns with other workers. Using tools from the optimal transport literature, I solve the model and structurally estimate it on Portuguese matched employer-employee data. The Portuguese labor market is characterized by an increase in the relative supply of high school graduates, an increasingly unbalanced distribution of high school graduates versus non-graduates across industries, and a decreasing high school wage premium between 1987 and 2017. Counterfactual exercises suggest that both changes in worker preferences and the increasing relative productivity of high school graduates over non-graduates act as a mitigating force on the decreasing high school wage premium, but do not fully compensate for high school graduates' rise in relative supply.

Keywords: Educational Changes, One-to-Many Matching, Structural Estimation

JEL Codes: C7, D2, J2

*Sciences Po, pauline.corblet@sciencespo.fr

I thank my advisor Alfred Galichon for his valuable guidance and support. My Ph.D is partly financed through the ERC grant EQUIPRICE. I am indebted to Francesco Franco who generously helped me obtain the data used in this project. I also thank Jean-Marc Robin, Moshe Buchinsky, Florian Oswald, Kerstin Holzheu, Zsofia Barany as well as all the participants to Sciences Po PhD and Equiprice seminars for their comments and feedback. All remaining errors are mine.

1 Introduction

What can sorting patterns and wages on the labor market tell us about underlying workers' and firms' preferences? Workers hired by a firm interact to produce output, and each employee contributes differently to total output depending on their characteristics (education level, experience, etc.). These characteristics are more or less substitutable, in the sense that the same level of output may be obtained from different combinations of workers characteristics. Workers' level of contribution also depends on firm's sector, or on the type of goods it produces. Besides, workers may derive different levels of satisfaction (that do not come from wage) from being employed at different firms. Workers with a given characteristic, and firms in a given sector both are in limited supply on the labor market. The structure of firm production and worker preferences, which shapes workers' and firms' demand, along with the availability of workers and firms, which forms their supply, determine two main outcomes on labor markets: sorting and wage. Sorting is a relative measure of match strength between workers and firms with given characteristics. Wages are the monetary transfers made from firms to workers in a match. Both worker and firm supply and demand are likely to evolve over time, and both sorting and wages evolve as a result. Apprehending the relation between supply and demand on the one hand and sorting and wages on the other is central to understanding labor markets, and has been the focus of a large economic literature. This paper builds a novel one-to-many matching model to capture the relation between supply and demand and labor market outcomes. I then use the model to explore the reasons behind the increasingly unbalanced sorting of workers and firms and the decreasing education wage premium in Portugal between 1987 and 2017.

I build a static one-to-many matching model with transferable utility in which workers and firms differ with respect to their observed characteristics, that are summed up by a multidimensional type, as well as a stochastic shock that accounts for unobserved heterogeneity. A single firm matches with several workers, who constitute a bundle that forms its workforce. The surplus created by the match depends on the firms' observable characteristics as well as the workforce's. Utility is transferable under the form of wages paid by the firm to the workers in its workforce. Firms seek to maximize total profit, which is additive in the difference of production and total wage bill, plus random shocks. Workers maximize their utility, which is additive in amenities, wage and a random shock. Amenities embody workers' inner preference for a given type of firms. At equilibrium, wages clear the market and each agent match with their best option given wages. The model is able to generate a rich distribution of wages that depend both on worker's and firm's observable characteristics, as well as on employed workforce. It also predicts equilibrium matching, which is the joint distribution of firms and workforces. Using both matching and wages, I am able to separately identify firm production from workers' amenity.

The framework offers more flexibility in estimation than classic supply and demand models developed in Katz and Murphy (1992) and Card and Lemieux (2001): it identifies worker preferences in addition to firm production, as well as varying parameters over time. This is because by explicitly modeling firms and workers match choices, I am able to use both observed matching and observed wages, which brings more power to identification. The model is fitted to the data by assuming parametric forms for firm production and workers amenities. I classify workers into two education levels, high school graduates and non graduates, and three age groups, young, middle-aged and senior. Firms are differentiated by their sector of activity. Following the literature, I choose a nested Constant Elasticity of Substitution (CES) function for production, with productivity parameters for each education level that vary between sectors. I assume worker preferences for firms to depend on worker's own age and education level, and firm sector. Equipped with model predictions for matching and wages, I am able to structurally estimate the model on matched employer-employee data. I estimate the model by maximum likelihood on the joint distribution of matching and wages, separately every three years.

I use the theoretical framework developed to study the Portuguese labor market between 1987 and 2017. I highlight three facts on the Portuguese labor market: first, the country operates a vast education expansion over the period, which translates in a dramatic increase in the relative supply of high school graduates to non-graduates on the labor market. Second, the high school wage premium decreases over the period. The high school wage premium is defined as the wage gap between workers who graduated from high school (including workers who pursued higher education), and those who did not. The decrease in wage premium is particularly stark among young workers, and co-occurs with an increase in the supply of young high school graduates compared to young non-graduates on the labor market. Third, I measure worker-firm sorting, which is defined as the relative number of high school graduates over non graduates in an age group employed in a given sector. The distribution of high school graduates versus non graduates across industry sectors becomes highly unbalanced, in favor of services, and transports and communications, who employ an increasing share of high school graduates. The former two facts implies relative supply of high school graduates over non-graduates has grown faster than firms relative demand for high school graduates over non-graduates. The latter suggests that sorting between workers and firms has evolved over the period: either because firms in services and transport and communications demand an increasing share of high school graduates, or because high school graduates' preference for these firms strengthens. In the first scenario the impact on high school wage premium in services and transport and communication should be positive, and in the second scenario it should be negative.

Portugal is a particularly relevant example of rapid supply and demand changes on the labor market: it entered the European Union in 1986, which fuelled its economy's transition from being dominated by manufacturing (50% of the labor force employed in 1987), to services (30% of the labor force employed in 2017). Meanwhile, only 10% of its employed labor force held a high school degree in 1987, a percentage that has risen to 50% in 2017. As a point of comparison, the percentage of high school graduates in the US workforce has gone from 75% to 90% over the same period[1]. The proportional increase of high school graduates in Portugal is more extensive and starts from a much lower presence of high school graduates on the labor market than in the US. In this respect, it is closer to the change in university graduates on the US labor market (from 20% to 35% over the same period). Graduating from high school has become much more common in Portugal over the last thirty years, but it is only in 2007 that high school graduates started representing the majority of young workers between 25 and 30. In 2017, 32% of the young workers between 25 and 30 still do not hold a high school degree. Meanwhile, university graduates in Portugal represented less than 3% of the employed labor force in 1987, and about 19% in 2017. Because the share of university graduates remain small for most of the period (it only reaches 10% in 2005), and because graduating from high school is still quite uncommon over most of the period I study, I consider a high school degree to be a differentiating signal in skill on the Portuguese labor market, much as a college degree is on the US labor market.

I find that relative demand for high school graduates from firms in the services, manufacturing and transport and communications sectors has increased dramatically over the period, starting in the early 2010s. This finding is in line with the skill-biased technological change hypothesis. I also find that young and middle-aged high school graduates' preference for these industries has declines over time, while their share in production increases compared to senior workers. Compared to the classic supply and demand framework, these observations offer two additional mechanisms whereby high school wages gaps stay positive when large number of high school educated workers enter the labor market. First, a decrease in workers' utility to work pressures wage upwards. Second, variation in young graduates' share

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

in production compared to more senior high school graduates increases firm demands for the former compared to the latter. I perform a number of counterfactual exercises to assess the separate actions of changes in workers demographics (both in education and age distribution), firm sector composition, firm demand through production parameters, and worker preferences, on sorting and wage premium. I find that changes in demographics are the main positive drivers of changes in sorting. Changes in industry composition, firm demand and worker preferences overall have a negative, but more modest, effect on sorting. Wage premia by age group and industry are negatively affected by changes in worker demography and industry composition, and positively affected by changes in firm demand. These suggest changes in relative productivity in favor of high school graduates has driven the high school wage premium up, but cannot compensate for the large increase in relative supply of graduates versus non-graduates.

Related literature. The theoretical tools developed in this paper belong to the matching literature started by Becker (1973). My model is a one-to-many extension to the seminal work of Choo and Siow (2006) in the one-to-one case. As in Dupuy and Galichon (2017) and Galichon and Salanie (2020), it explicitly borrows tools from the optimal transport literature to introduce unobserved heterogeneity in the form of random utility, and it relies on Gretsky et al. (1992) to show equilibrium existence. This framework differs from the Sattinger model (Sattinger (1979), Sattinger (1993)) that assumes no unobserved heterogeneity and rests on the the firm's production function's supermodularity to find the optimal assignment of workers to firms. Because static random utility models (including mine) do not follow agents over time, they do not identify unobserved characteristics' contribution to match surplus in the fashion of Abowd et al. (1999), Bonhomme et al. (2019) and Bonhomme (2021), and instead focus on match formation based on observable surplus. Fox (2009) discusses non parametric identification in matching games and Fox et al. (2018) show that unobserved heterogeneity distribution can be recovered in matching games in which unmatched agents are observed. My work is also related to seminal work by Kelso and Crawford (1982) and Bikhchandani and Ostroy (2002) on one-to-many matching, and more recent work by Che et al. (2019) on one-to-many matching with non-transferable utility and Azevedo and Hatfield (2018) on one-to-many matching with transferable utility. They both show existence of equilibrium for a large class of firm preferences, under a large market assumption, an assumption I also use in this paper.

The model I develop features sorting between multidimensional types and as such is also related to Choné and Kramarz (2021), Lindenlaub (2017) and Postel-Vinay and Lise (2015). However, it is only remotely related to the search literature to which the latter paper belongs, as it focuses on relative supply and demand instead of search frictions. While the search literature often relies on Nash bargaining mechanisms, as in Shimer and Smith (2000) or Cahuc et al. (2006), the present model uses wage posting, as the competitive equilibrium in the model rests on wages that clear the labor market. Also related to this model and its application is the Roy model developed by Hsieh et al. (2019) to quantify the productivity gains of weakening discrimination barriers to women's and black men's entry on the labor market in the US. There exists a large and extensive literature on the education wage premium, mostly focused on the college wage premium in the US. Seminal work by Katz and Murphy (1992) shows that the increasing supply of college graduates in the 1970s and 1980s is absorbed on the US labor market by an increased demand for these workers from firms. Card and Lemieux (2001) carry out a similar analysis that further differentiates workers by age, and show that young college graduates are the first to benefit from the slowdown in education attainment in the 1980s. Goldin and Katz (2008) and Autor et al. (2020), among others, relate changes in the US wage structure to be race between education and technology, by which skill biased technological change favors college graduates. Skill-biased technological change (SBTC) origins in the development of new technologies, in particular computers (Autor et al. (1998), Autor et al. (2003)). However, if the SBTC hypothesis has proven a powerful explanation for the quick increase in graduate wage premium of the 1970s and 1980s, it is less clear if it can rationalize the subsequent slow down of both graduate wage premium and graduate supply in the 1990s, when the use of computers became prevalent (Card and DiNardo (2002)). The recent stagnation of the college wage premium in the US is also documented in a number of papers, and several explanations have been put forward: Beaudry et al. (2015) argue that the demand for cognitive skills has decreased since the early 2000s, pushing graduate workers down the job ladder. Valletta (2016) also emphasizes the role of job market polarization, i.e. the shift away from middle-skilled occupations, on college graduates' wages (as opposed to postgraduates). On the contrary, Blair and Deming (2020) examine job vacancy data and find that demand for skills has increased since the Great Recession. They explain the stagnating graduate wage premium by an increase in the supply of new graduates after 2008. They are backed by Hershbein and Kahn (2018) who show that the Great Recession has accelerated skill-biased technological change. In Portugal, changes in the wage structure are documented by Cardoso (2004), Centeno and Novo (2014) Almeida et al. (2017). To the best of my knowledge, I am the first to analyze the implications of worker and firm sorting on the education wage premium.

Outline Section 3 describes the evolution of the Portuguese high school wage premium between 1987 and 2017. Section 2 describes the one-to-many matching model. Section 4 discusses the model's identification and estimation, and section 5 presents estimation results. Section 6 compares the results with the simple model of Card and Lemieux (2001). Section 7 concludes.

2 Model

Recent administrative matched employer-employee datasets hold much more information than workers' characteristics and wage. They also inform on firms and on matching, i.e. the joint distribution of workers and firms. Besides matching, we also observe the transfers between agents, in the form of wage. Relying on these datasets enables to build a rich supply and demand framework to understand the race between education and technology. I build a one-to-many matching model where a single firm matches with several workers, who interact within the firm to produce output. Workers are compensated through wage, and hold specific preferences for different types of firms. Workers may also be unemployed. Firms maximize their profit, given their production function that is specific to their type and market clearing wage. Both worker anf firm types are observed, and possibly multidimensional. The model is an extension of Choo and Siow (2006) to a one-to-many framework, and existence of equilibrium rests on a large market assumption, as in Azevedo and Hatfield (2018) and Galichon and Salanie (2020). I model unobserved heterogeneity in the form of additive random utility. The social planner problem the rewrites as a regularized optimal transport problem (Galichon (2016)) and I am able to derive closed-form solutions for predicted matching and wage.

2.1 Set Up

The labor market is two-sided, with workers and firms on each side. There is a continuum of workers $i \in I$. Each worker has a type $x \in \mathcal{X}$. Types are discrete and possibly multidimensional. There is a mass n_x of workers of type x, and a finite number of types: $\#\mathcal{X} = X$. On the other side of the market, there is a large number of firms $j \in J$. Each firm has a type $y \in \mathcal{Y}$. As for workers, firm types are discrete and possibly multidimensional. There is a mass m_y of firms of type y, and a finite number of types: $\#\mathcal{Y} = Y$.

Each firm matches with a non negative number of worker of each type, while each worker matches with a single firm. Let k_x be the number of type x workers a firm is matched with. The model is scaled by scaling factor F, so that the number of type x workers on the market is Fn_x. Therefore k_x must be comprised between 0 (a firm cannot hire a negative number of workers), and Fn_x. Vector k is the workforce employed by the firm. It is akin to a bundle of workers of each type:

$$k = (k_1, \ldots, k_X) \in [0, Fn_1] \times [0, Fn_X].$$

Type x worker's utility for being employed at type y firm within workforce k is u_{xyk}. It is additive in a level of amenity α that depends both on worker and firm type, as well as workforce, and in wage w paid by the firm to the worker. Wage w_{xyk} is also allowed to depend on worker type, firm type and workforce.

$$u_{xyk} = \alpha_{xyk} + w_{xyk}.$$

Every worker also has the option to remain unemployed and obtain $u_{x0} = 0$.

Similarly, the firm profit is additive in production γ and minus total wage bill paid to its workforce.

$$v_{yk} = \gamma_{yk} - \sum_{x=1}^{X} k_x w_{xyk}.$$

Both amenity α_{xyk} and γ_{xyk} are functions of x, y, k and take their value in $\mathbb{R}$. The total surplus from a match between a firm and a workforce is the sum of workers' utilities and firm's profit

$$\Phi_{yk} = \sum_{x=1}^{X} k_x \alpha_{xyk} + \gamma_{yk} \tag{1}$$

where wages have canceled out because they are modelled as perfectly transferable utility.

Some firm and workers characteristics that play a role in match formation are unobserved, and therefore are not accounted for in x or y. There exists a large literature that deals with unobserved heterogeneity, and I build on a large subset (Choo and Siow (2006), Dupuy and Galichon (2014)) that uses additive random shocks to model it. I further assume a logit framework for the model by restraining the distribution of shocks to belong the extreme value class, although as shown in Galichon and Salanie (2020) in the one-to-one case identification is possible with a general class of distributions.

Worker i experiences stochastic shock $(\epsilon_{iyk})_{y,k}$ in addition to her systematic utility:

$$u_{x_i yk} + \xi \epsilon_{iyk}.$$

Similarly firm j experiences stochastic shock $(\eta_{jk})_k$ in addition to its systematic production:

$$v_{y_j k} + \xi \eta_{jk}.$$

where ξ is a scaling factor. I impose the following independence conditions on stochastic shocks:

Assumption 1. *Stochastic shocks satisfy the following:*

(i) For each pair of two workers i and i', ϵ_{iyk} and $\epsilon_{i'yk}$ 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 $\alpha_{x_i yk}$, η_{jk} is independent of γ_{yk}.

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

A market is characterized by exogenous distributions of worker and firm types $(n_x)_{x\in\mathcal{X}}$ and $(m_y)_{y\in\mathcal{Y}}$, as well as amenity functions $(\alpha_{xy})_{x\in\mathcal{X},y\in\mathcal{Y}}$, production functions $(\gamma_y)_{y\in\mathcal{Y}}$, and a draw of stochastic shocks ϵ and η. In the next subsection, I describe workers and firms choices and the resulting competitive equilibrium

2.2 Competitive Equilibrium

Next I define workers and firms expected utility and profit from choosing their best employer or workforce, given wages.

Definition 1. *Type x worker's expected indirect utility G_x as a function of u and type y firm's expected indirect utility H_y as a function of v are*

$$G_x(u_x) = \mathbb{E}\left[\max_{y,k}\left\{u_{xyk} + \xi\epsilon_{yk}, \xi\epsilon_0\right\}\right] \quad \text{and} \quad H_y(v_y) = \mathbb{E}\left[\max_{k}\left\{v_{yk} + \xi\eta_k\right\}\right].$$

Under assumption 1, expected utilities rewrite in closed form.

Proposition 1. *Under assumption 1, expected indirect utilities write*

$$G_x(u_x) = \xi \log\left(1 + \sum_y \sum_k \exp\left(\frac{u_{xyk}}{\xi}\right)\right) \quad \text{and} \quad H_y(v_y) = \xi \log \sum_k \exp\left(\frac{v_{yk}}{\xi}\right)$$

where $\sum_k = \sum_{k_1} \dots \sum_{k_X}$

Proof. In appendix B. □

The equilibrium on a market is found when supply from workers meets demand from firms. Supply and demand are defined as follows:

Definition 2. *Type x worker's supply is a vector $(S^x_{yk})_{yk,0}$ where S^x_{yk} is the mass of type x workers willing to match with type y firm and workforce k and S^x_0 is the mass of type x workers willing to remain unmatched.*

Type y firm's demand is a vector $(D^y_k)_k$ where D^y_k is the mass of type y firms willing to match with workforce k.

I model unemployment through S^x_0, which is determined at equilibrium. I assume no counterpart on the firm side: all firms must be matched to a given workforce.

Assumption on stochastic shocks lets us express supply from worker and demand from firms in logit form.

Proposition 2. *Under assumption (1), the mass of type x workers willing to supply type y firms in workforce k is*

$$S^x_{yk} = n_x \frac{\exp\left(u_{xyk}\right)}{1 + \sum_{y,k} \exp\left(u_{xyk}\right)} \tag{2}$$

The mass of type y firms who demand workforce k is

$$D^y_k = m_y \frac{\exp\left(v_{yk}\right)}{\sum_{y,k} \exp\left(v_{yk}\right)} \tag{3}$$

Proof. In appendix B. □

Note that supply S and demand D both depend on wage schedule $w = \left(w_{xyk}\right)_{x,y,k}$. Because both workers and firms care not only about the other side's type, but also about the workforce they work with both in the systematic and stochastic parts of their utility or profit, wages also depend on workforce k. Therefore, two type x workers employed in two firms of same type y but who hire different workforce k and k' do not receive the same wage, as $w_{xyk} \neq w_{xyk'}$ in general. The model is able to generate heterogeneity in wage depending on firm size and workforce composition.

In the context of one-to-many matching, supply S and demand D are measured in different 'units': if a firm can match with several workers types, workers can only match with one firm type. Excess demand Z defined below gives the equivalence between worker and firm units.

Definition 3. *Given types x, y and workforce mass k, excess demand is defined as*

$$Z_{xyk}(w) = k_x D_k^y - S_{yk}^x.$$

A competitive equilibrium is reached on the market when supply and demand are feasible, matching is incentive compatible, and excess demand is zero. The first two conditions are automatically filled as a byproduct of the definition of supply and demand: in proposition 2, workers and firms choose their optimal option. As a result, matching is incentive compatible, and supply and demand are feasible:

$$\sum_{y,k} S_{yk}^x + S_0^x = n_x \quad \text{and} \quad \sum D_k^y = m_y$$

Definition 4. *An equilibrium outcome (S, D, w) satisfies $\forall x, y, k$: $Z_{xyk}(w) = 0$*

The existence a competitive equilibrium rests on the fact that there are large numbers of agents on the market. To show existence, I follow a proof technique introduced in the continuum assignment problem by Gretsky et al. (1992), and already used for one-to-one matching markets by Galichon and Salanie (2020) The reasoning is also very close to Azevedo and Hatfield (2018)'s proof for competitive equilibrium existence in a large economy on a market of buyers and sellers with a finite set of possible trades. Bikhchandani and Ostroy (2002) explore a similar assignment problem but do not assume marge markets and work without heterogeneous shocks.

I prove existence of equilibrium in two steps. First, I show that the competitive equilibrium reframes as an optimization problem on total welfare. Second, I show this problem is the dual of the social planner problem, who maximizes total surplus under feasibility conditions. The social planner problem maximizes a continuous and strictly concave function over a compact space. As such, a unique solution exists.

Theorem 1. *Equilibrium payoffs obtain as solutions to the following problem:*

$$\begin{aligned} &\inf_{u,v} \sum_x n_x G_x(u_x) + \sum_y m_y H_y(v_y) \\ &s.t \sum_x k_x u_{xyk} + v_{yk} = \Phi_{yk} \quad \forall k, y \end{aligned} \tag{4}$$

Proof. In appendix B ∎

Theorem 2. *Equilibrium matching* $\mu_{yk} = \frac{S^x_{yk}}{k_x} = D^y_k$ *and equilibrium* S^x_0 *obtain as solution to the social planner problem:*

$$\begin{aligned} \max_{\mu, S_0} & \sum_y \sum_k \Phi_{yk}\mu_{yk} + \xi \mathcal{E}(\mu, n, m) \\ s.t & \sum_y \sum_k k_x \mu_{yk} + S^x_0 = n_x \\ & \sum_k \mu_{yk} = m_y \end{aligned} \tag{5}$$

where $\mathcal{E}(\mu, n, m)$ *is equal to*

$$\begin{aligned} \mathcal{E}(\mu, n, m) = & -\sum_x n_x \sum_y \sum_k \frac{k_x \mu_{yk}}{n_x} \log \frac{k_x \mu_{yk}}{n_x} - \sum_x n_x \frac{S^x_0}{n_x} \log \frac{S^x_0}{n_x} \\ & - \sum_y m_y \sum_k \frac{\mu_{yk}}{m_y} \log \frac{\mu_{yk}}{m_y} \end{aligned}$$

The solution to (5) *exists and is unique.*

Proof. In appendix B. □

Theorem 2 shows that equilibrium matching can be obtained by solving a penalized social planner problem, where the objective function is the difference between total expected surplus and an entropy term due to unobserved heterogeneity. It is reminiscent of the discrete regularized optimal transport problem (Galichon (2016)). However it differs from the usual transport problem in two important ways: first workers are allowed to remain unmatched through S^x_0, and second, the first marginal condition $\sum_y \sum_k k_x \mu_{yk} + S^x_0 = n_x$ is not a condition on the marginal distribution of k, which is endogeneous, but on the marginal distribution of worker types.

Solving for problem (2) yields the following expressions for equilibrium matching μ, unemployment S^x_0 and wages w.

Proposition 3. *Equilibrium matching solves*

$$\begin{aligned} \log \mu_{yk} &= \frac{\Phi_{yk} - \sum_x k_x U_x - V_y + \sum_x k_x \log \frac{n_x}{k_x} + \log m_y}{\xi (1 + \sum_x k_x)} \\ \log S^x_0 &= \frac{-U_x + \log n_x}{\xi} \end{aligned} \tag{6}$$

Equilibrium wages write

$$
\begin{aligned}
w_{xyk} = & \frac{\gamma_{yk} - \alpha_{xyk} + U_x - V_y + \log m_y - \log \frac{n_x}{k_x}}{\xi(1 + \sum_x k_x)} \\
& + \frac{\sum_{x' \neq x} k_{x'} \left((\alpha_{x'yk} - \alpha_{xyk}) - (U_{x'} - U_x) + \log \frac{n_{x'} k_x}{n_x k_{x'}} \right)}{\xi(1 + \sum_x k_x)}
\end{aligned}
\tag{7}
$$

Where U_x, V_y solve

$$
\begin{cases}
\sum_{y,k} k_x \exp\left(\frac{\Phi_{yk} - \sum_x k_x U_x - V_y + \sum_x k_x \log\left(\frac{k_x}{n_x}\right) + \log m_y}{\xi(1+\sum_x k_x)} \right) + \exp\left(\frac{-U_x + \log n_x}{\xi} \right) = n_x \\
\sum_k \exp\left(\frac{\Phi_{yk} - \sum_x k_x U_x - V_y + \sum_x k_x \log\left(\frac{k_x}{n_x}\right) + \log m_y}{\xi(1+\sum_x k_x)} \right) = m_y
\end{cases}
\tag{8}
$$

Proof. In appendix B □

In practise, equilibrium μ, S_0^x and w are computed by solving for equations (8) using the Sinkhorn algorithm, also called IPFP, that has been developed in the optimal transport literature (among others). In the one-to-many case, U_x and V_y can be solved for by coordinate update in the same spirit as Sinkhorn.

2.3 Links with search and matching models in the literaure

The model I develop is akin to Choo and Siow (2006)'s in a one-to-many instead of a one-to-one setting. One can view the space of workforces, instead of workers, as a side of the market, with firms on the other side. It is particularly striking that just like in Choo and Siow (2006), both equilibrium matching and wage are weighted by the number of individuals in the match $1 + \sum_x k_x$. In this representation, the model almost reduces to the one-to-one framework, but for the specific shape of marginal conditions in (8), that links the matching over workforces and firms back to the number of workers of each type. Another difference with Choo and Siow (2006), Dupuy and Galichon (2017) and other frameworks that use the IPFP algorithm in their framework is that expected indirect surpluses U and V cannot be explicitly expressed through equations (8) because the size of every match is endogenous. I observe transfers as wages and can leverage them to split total match surplus between workers and firms, in the spirit of Dupuy and Galichon (2017)

The model also features wage posting. In the decentralized equilibrium, firms choose among workforces and associated wages given their draw of random shock η, while workers choose among firm types, workforces and wages given their draw of ϵ. A salient feature of the model is that it generates wage dispersion for a given worker and firm type, based on the workforce hired by the firm. All other things equal, wage in increasing in the number of workers hired by the firm. This is reminiscent of search models such as Burdett and Mortensen (1998), that there is no search in the model presented here.

Finally, the model is closer to Katz and Murphy (1992) and Card and Lemieux (2001) than it may appear at first sight. To see this, consider two workforces k and k', where $k'_x = k_x$, expect for $k'_{\bar{x}} = k_{\bar{x}} + t$, i.e. there is t more worker of type $\bar{x}$ hired in workforce k'. Then firm production and type $\bar{x}$ worker's wage satisfy:

$$\gamma_{yk} - \gamma_{yk'} = \left(1 + \sum_x k_x\right) w_{\bar{x}yk} - \left(1 + \sum_x k'_x\right) w_{\bar{x}yk'}.$$

At the limit, when t tends to zero (if the extra worker workers very few hours for instance), we obtain the same intuition as with the representative firm that marginal change in wage is equal to marginal change in production (divided by the number of agents):

$$\frac{\partial \gamma_{yk}}{\partial k_x} = \left(1 + \sum_x k_{\bar{x}}\right) \frac{\partial w_{\bar{x}yk}}{\partial k_{\bar{x}}}.$$

Hence any change in workers' $\bar{x}$ is proportional to their marginal productivity, although its impact is mitigated by total number of workers hired by the firm.

3 Empirical Evidence

3.1 Data Description

The *Quadros de Pessoal* dataset offers an exhaustive snapshot of the Portuguese labor market every year from 1987 to 2017. It covers all employees in the private sector (except domestic workers), and provides information on their age and highest degree obtained, as well as their monthly wage and hours worked. To compute the high school wage premium by age, I part the worker population into two groups: those who did not graduate from high school, and those who did. I also categorize workers into three age groups: young workers (from 16 to 35 years old), middle aged workers (from 36 to 50 years old), and senior workers (from 51 to 68 years old). I only consider full time employees, that is, workers that are neither part time workers (approximately 10% of the observations) nor self-employed, in unpaid family care, or in other forms of employment (less than 1% of the observations).

I compute real hourly wage as the ratio of monthly wage over monthly hours, controlling for inflation and clean out the lowest 1% and highest 99% hourly wage percentiles. Firms belong to either five sectors, or industries: primary industries (agriculture, mining, energy, construction), manufacturing, retail and hospitality, services, transport and communications.

To account for unemployment, I use public yearly unemployment figures by education level and age group provided by INE[2]. Information on unemployment is missing between 1987 and 1991, hence I assume the unemployment rate in these years is the same as in 1992. I compute the number of unemployed workers each year by education level and age group by combining unemployment rates and the number of observed employed workers in *Quadros de Pessoal*. In what follows, active worker refers to workers either employed or unemployed.

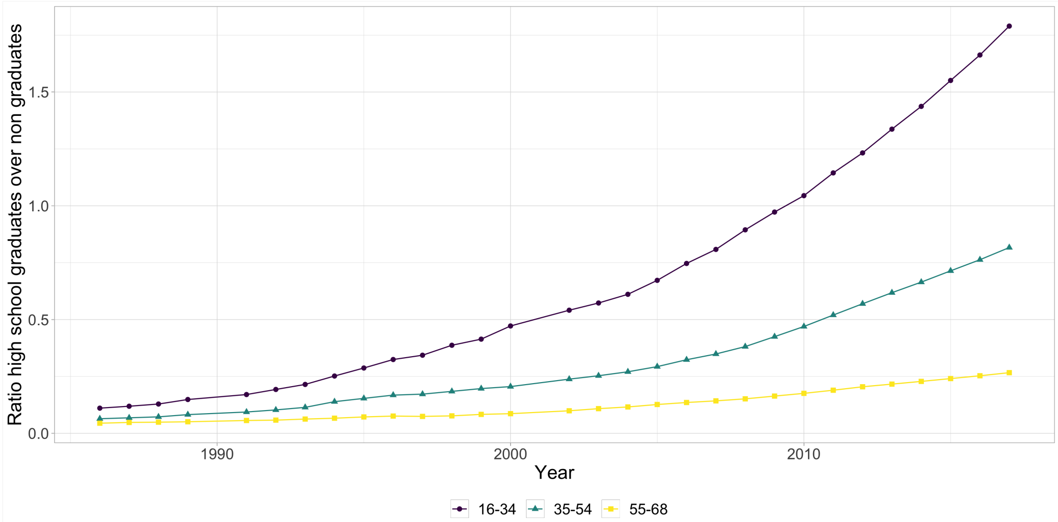
3.2 Empirical facts

The Portuguese labor market is characterized by three facts between 1987 and 1997. The first is the dramatic increase in the number of high school educated workers, compared to the number of workers who did not go to high school. The second is the decrease in high school wage premium, i.e. the wage gap between high school graduates and non graduates. The third is the change in sorting between education level on the worker side, and industry on the firm side: sorting intensity between high school graduates and specific industries rises over the period. Each of these three facts are detailed below.

Fact 1: Education supply. Relative supply of high school graduates versus non graduates rises dramatically over the period, as evidenced by figure 1. Relative supply is measured as the ratio of number of active high school graduates over number of active non high school graduates by age group in each year. Because high school enrolment grows every year, young workers are most impacted by this growth, and their relative supply goes from .12 to 1.79 on figure 1, meaning high school graduates have grown to be about eight times less numerous to almost twice as numerous as non graduates between 1987 and 2017.

[2] Found on their website

Figure 1: High school graduates versus non graduates relative supply, by age group



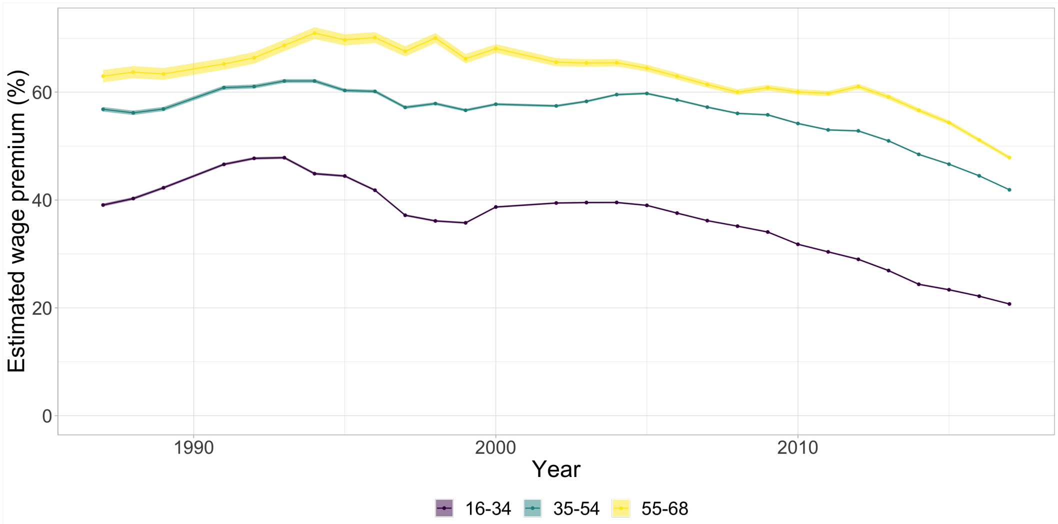
Fact 2: Wage premia by age group. The second fact that characterizes the Portuguese labor market is the decrease in high school wage premium. To compute high school wage premium by age group, I estimate the following by OLS:

$$\log w_{ijt} = \sum_{a_i \in \{y,m,s\}} \mathbb{1}_{[\text{HS graduate}_i]} \beta_{at} + g_i + r_{jt} + d_{jt} + u_{ijt} \tag{9}$$

where each individual i working in firm j at time t earns wage w_{ijt}. $\mathbb{1}_{[\text{HS graduate}_i]}$ is 1 if i graduated from high school, and 0 otherwise. a_i is individual i's age group: either y, m or s. g_i, r_{jt} and d_{jt} are gender, region and industry fixed effects. β_{at} is the yearly high school wage premium, differentiated by age group: it measures how much more in percentage a high school graduate earns compared to a non high school graduate. To let fixed effect vary over time, I estimate (9) separately every year.

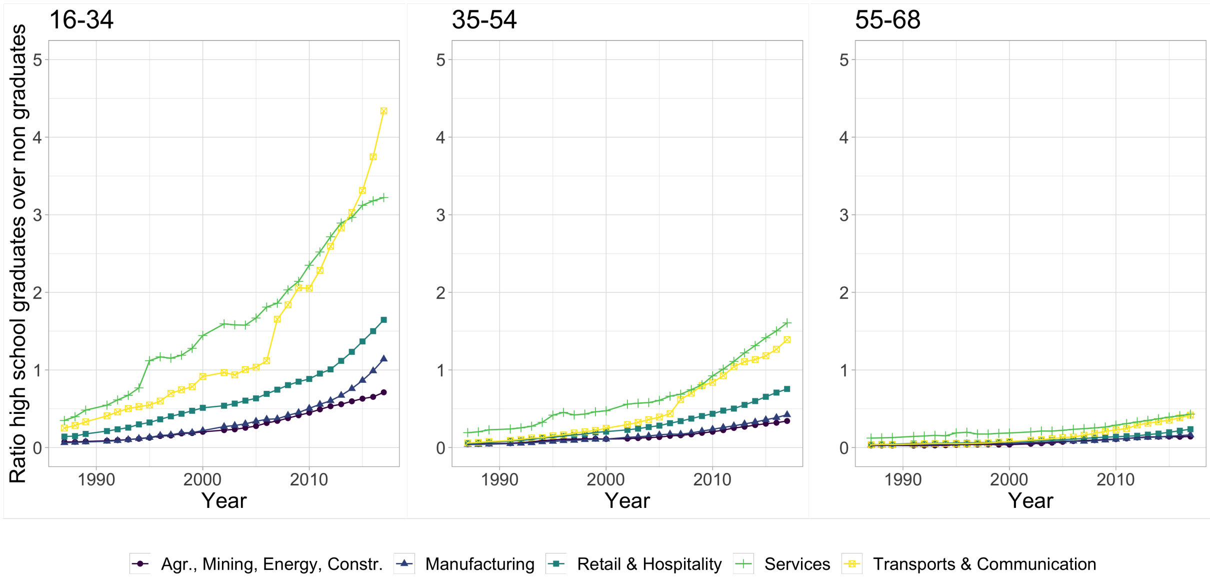
Figure 2 shows the change in estimated high school wage premium over time for each age group, along with 5% confidence intervals. It shows high school wage premia differs by age group: it is much higher (between 60% and 80%) for senior workers than for younger workers (between 40% and 20%). It also shows that the wage premium decreases for all age groups between 1987 and 2017. But the extent of the decrease is different depending on age: senior workers lose only about 17 percentage points (p.p) in high school wage premium over the period, while young workers lose almost 50p.p and middle ages workers lose slightly less than 30p.p.

Figure 2: Estimated high school graduate wage premium over non graduates of same age



Fact 3: Sorting between education levels and sector. Sorting between education level and industry is measured by age group as the ratio of number of employed high school graduates to employed non-graduates in a sector. Sorting is said to be stronger between high school graduates and sector A than sector B, if this ratio is larger in sector A than in sector B. Plotting sorting ratios by sector over time reveals stark differences by industry, as shown in figure 3. Most notably, the services and transport and communications industries hire young high school graduates over non graduates at a higher rate than the change in overall relative supply. As shown in fact 1, relative supply goes from .11 to 1.79 over the period, while the sorting ratio in these industries reaches 3.22 and 4.34 in 2017. Services and transports and communications also hire proportionally more middle-aged workers, with a ratio of 1.61 and 1.39 in 2017, compared to a relative supply ratio of .82.

Figure 3: High school - sector sorting, by age group



Summary. The relative supply of high school graduates over non graduates rises for all age groups, and in particular among young workers. Meanwhile, the high school wage premium decreases in Portugal between 1987 and 2017. Its decline is particularly strong for young workers, between 16 and 34 years old. The rise in relative supply is not absorbed equally by all sectors: services and transports and communications hire proportionally more young and middle-aged high school graduates than other sectors. This is indicative of strong sorting between these workers and the services and transports and communications industry.

Portugal is unique in that it has known a dramatic education expansion, going from 10% of high school graduates in the labor force in 1987 to about 50% in 2017. It has also known deep changes in how workers sort with firms based on education level, age group, and the firm sector, as evidenced in fact 2. As such, it is an ideal laboratory to understand how sorting between workers and firms drives the high school wage premium overt time. Changes in sorting can be caused either by an increase in relative productivity of high school graduates in some industries, a change in preferences of young high graduates, or changes in substitution patterns among education levels or age groups. Meanwhile, the increase in relative supply of high school graduates likely drives the wage premium down. In the next section, I build a framework to quantify these changes, untangle the effect of changes in relative supply from changes in firm production and worker preferences, and evaluate their impact on sorting and wage.

4 Identification and Estimation

The model's predictions on matching (6) and wage (7) allow to separately identify amenity and productivity functions $(\alpha_{xy})_{xy}$ and $(\gamma_y)_y$. This would not be true if we observed only matching, as α and γ appear together in the matching prediction, and only total surplus can be identified from this equation. If only wages were observed, the same problem arises and only the difference between firm production and worker amenities is identified. In this case one must assume that amenities are zero in order to identify production.

In any given period t, I aim at parametrically estimate α^t and γ^t. All amenity and production parameters are allowed to vary with time, and in what follows I drop the superscript t to ease the exposition. I assume $N = 6$ worker types that are the combination of two education levels, and three age groups. The education levels are high school graduates H and non graduates L, and the age groups are young y (below 35), middle-aged m (between 35 and 54), and senior s (above 55). Let $e(x)$, $a(x)$ be type x's education level and age group. Firm workforce k is composed of the numbers of each worker type employed

$$k = \left(k_{H,y}, k_{H,m}, k_{H,s}, k_{L,y}, k_{L,m}, k_{L,s}\right).$$

Employed number of worker k_x is directly observed in the data and defined as total number of hours worked monthly by workers of type x hired by the firm, divided by 174, the monthly hours equivalent of a 40 hours week. Hence each k_x counts the full-time equivalent of the number of type x workers employed by the firm. This measure is not necessarily an integer, as part-time workers would count as fractions of the full-time equivalent. Type y firm produces according to a nested Constant Elasticity of Substitution (CES) production function with different parameters depending on its type y:

$$\gamma_y(k) = \left[\left(\theta_H^y H(k)\right)^{\frac{\sigma-1}{\sigma}} + \left(\theta_L^y L(k)\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} - \sum_x \mathbb{1}_{[k_x>0]} \frac{\nu_y}{n_x}$$

where aggregates $H(t)$ and $L(t)$ are:

$$H(k) = \left[\sum_{a\in\{y,m,s\}} \lambda_{a,H} k_{H,a}^{\frac{\tau^H-1}{\tau^H}}\right]^{\frac{\tau^H}{\tau^H-1}} \quad \text{and} \quad L(k) = \left[\sum_{a\in\{y,m,s\}} \lambda_{a,L} k_{L,a}^{\frac{\tau^L-1}{\tau^L}}\right]^{\frac{\tau^L}{\tau^L-1}}.$$

Production γ^y's outer nest involves three parameters: σ, θ_H^y, θ_L^y and two aggregate inputs $H(k)$ and $L(k)$. $\sigma \in (0, \infty)$ is the elasticity of substitution between education levels, it is greater than one if high school graduates and non graduates are gross substitutes, and smaller than one if they are gross complements. σ is assumed to be the same across firm types. θ_H^y, $\theta_L^y \in [0, \infty)$ are graduates and non graduate's productivity parameters. Both parameters may vary by firm type y. In addition to their CES production function, firms experience friction $\frac{\nu_y}{n_x}$ if they employ workers of type x. The rationale is that if there are few workers of type x, then it is costly for the firm to find and hire them. ν_y measures this cost by sector y.

Aggregate labor inputs $H(k)$ and $L(k)$ form the production function's inner nest. They each depend on four parameters: three age productivity parameters each: $\lambda_{a,H}^{y,t}$ and $\lambda_{a,L}^{y,t} \in [0, \infty)$ and one elasticity of substitution between age levels each: τ^H and $\tau^L \in (0, \infty)$. Elasticities vary by education level but are the same across firm types, while age productivity vary with firm type y.

The production function is close to the one used by Katz and Murphy (1992), and Card and Lemieux (2001): it assumes imperfect substitution and varying productivity in the tasks performed by different education levels and age categories. Capital is not included as an input, but may impact productivity parameters through firm type: if two firm types use different levels of capital in relation to education levels, it is reflected in the levels of θ_H^y and θ_L^y. Unbiased technological change that increases all workers productivity results in an increase in both θ_H^y and θ_L^y. Technological change may be biased towards an education level if its productivity increases faster than the other's. This production function also allows more flexibility than Card and Lemieux (2001) by letting elasticities of substitution and age productivity vary in time.

Production assumes constant returns to scale. Note that it is homogeneous of degree one, and therefore two functions parametrized with θ and λ or $c \times \theta$ and $\frac{\lambda}{c}$ are equivalent. To distinguish between these versions, I impose normalization condition:

$$\sum_a \lambda_{a,H}^y = \sum_a \lambda_{a,L}^y = 1 \quad \forall y \tag{10}$$

I assume worker amenities are constant in k:

$$\alpha_{xy}(k) = \beta_x^y \tag{11}$$

β_x^y reflects type x worker preferences for type y firms over other firm types. In particular I assume workers are indifferent to workforce size.

Given these functional forms, I am looking to estimate in every period t parameters $\left(\lambda_{a,H}^y\right)_a$, $\left(\lambda_{a,L}^y\right)_a$, $(\theta_H^y)_y$, $(\theta_L^y)_y$, $(\beta_x^y)_{x,y}$, τ_H, τ_L and σ. To this aim I use a maximum likelihood method, which I describe in what follows.

The model predicts matching μ_{yk} as a joint distribution on firms and workforces, which can be compared to observed matching $\tilde{\mu}_{yk}$, which is simply the number of firms matched with workforces k in the data. Let also $\tilde{S}_0^x$ be the number of unemployed worker of type x. Let $\tilde{w}_{ij}$ be the observed wage of worker i employed by firm j. Observed wage $\tilde{w}_{ij}$ is assumed to be a noisy measure of predicted wage $w_{x_i y_j k_j}$ where k_j is the entire workforce employed at firm j. In other words:

$$\tilde{w}_{ij} = w_{x_i y_j k_j} + \nu_{ij} \text{ where } \nu_{ij} \sim \mathcal{N}(0, s^2) \text{ iid} \tag{12}$$

Where ν_{ij} is a centered measurement error of variance s^2. Under assumption (12), observed average wage $\tilde{W}_{xyk}$ for type x workers hired by firm y in workforce k is distributed as

$$\tilde{W}_{xyk} = \frac{1}{\tilde{K}_{xyk}} \sum_{\substack{i:x_i=x \\ j:y_j=y}} w_{x_i y_j k_j} \sim \mathcal{N}\left(0, \frac{s^2}{\tilde{K}_{xyk}}\right) \text{ iid} \tag{13}$$

where $\tilde{K}_{xyk}$ is the total number of type x workers hired by firm y in workforce k in the data: $\tilde{K}_{xyk} = k_x \tilde{\mu}_{yk}$. Because there is a very large number of observed wages in the data (as many as there are workers), I choose to work with observed average wages by worker type, firm type and workforce in the likelihood estimation. This reduces the likelihood function complexity but does not limit estimation: the model parameters as well as variance s^2 can still be recovered from log likelihood maximization.

Let $\mu_{yk}(\Gamma, \beta, n, m)$ and $w_{xyk}(\Gamma, \beta, n, m)$ be the matching and wage predicted by the model, given parameters $\Gamma = ((\theta_H^y)_y, (\theta_L^y)_y, (\lambda_{H,a})_a, (\lambda_{L,a})_a, \tau_H, \tau_L, \sigma)$, β, and worker and firm type distributions $n = (n_x)_x$, $m = (m_y)_y$. The log likelihood of observing pair $(x, y, k, \tilde{W})$ is then

$$\begin{aligned} &k_x \tilde{\mu}_{yk} \log \mu_{yk}(\Gamma, \beta, n, m) \\ &\qquad - \tilde{K}_{xyk} \frac{(\tilde{W}_{xyk} - w_{xyk}(\Gamma, \beta, n, m))^2}{2s^2} - \frac{1}{2} \log \left(\frac{s^2}{\tilde{K}_{xyk}} \right) \end{aligned}$$

Meanwhile, the log likelihood of observing an unemployed worker of type x is

$$\tilde{S}_0^x \log S_0^x(\Gamma, \beta, n, m, s^2)$$

The log likelihood method therefore solves

$$
\begin{aligned}
&\max_{\Gamma,\beta,s^2} l(\Gamma,\beta,n,m,s^2) \\
&= \max_{\Gamma,\beta,s^2} \sum_x \sum_{y,k} k_x \tilde{\mu}_{yk} \log \mu_{yk}(\Gamma,\beta,n,m,s^2) + \sum_x \tilde{S}_0^x \log S_0^x(\Gamma,\beta,n,m,s^2) \\
&\quad - \sum_x \sum_{y,k} \tilde{K}_{xyk} \frac{(\tilde{W}_{xyk} - w_{xyk}(\Gamma,\beta,n,m,s^2))^2}{2s^2} - \frac{1}{2}\log\left(\frac{s^2}{\tilde{K}_{xyk}}\right)
\end{aligned} \tag{14}
$$

I run log likelihood estimation on ten separate three year periods between 1987 and 2017[3]. Years in each period are pooled. In each period, I observe number of workers and firms $(\tilde{n}_x)_x$ and $(\tilde{m}_y)_y$ directly in the data. I normalize without loss of generality the total mass of firms in each period to 1, so that scaling factor F is $\sum_y \tilde{m}_y$, and input $n_x = \frac{\tilde{n}_x}{F}$ and $m_y = \frac{\tilde{m}_y}{F}$ to likelihood estimation.

I solve numerically for problem (14) using a nested method: in the inner loop, $\mu(\theta,\lambda,\tau,\sigma,\beta)$, $S_0^x(\theta,\lambda,\tau,\sigma,\beta)$ and $w_{xyk}(\theta,\lambda,\tau,\sigma,\beta)$ are computed according to (6) and (7). Scaling factor ξ is set to 1. In the outer loop, I update $(\theta,\lambda,\tau,\sigma,\beta)$ using Adam, a gradient descent method with momentum (Goodfellow et al. (2016), Kingma and Ba (2017)). Variance s^2 is obtained in the outer loop through first order condition:

$$
s^2 = \frac{1}{W} \sum_x \sum_{y,k} \tilde{K}_{xyk} \left(\tilde{W}_{xyk} - w_{xyk}(\Gamma,\beta,n,m)\right)^2
$$

More details on estimation can be found in appendix C.

5 Results

5.1 Parameters estimates

Estimates for ratio $\frac{\theta_H^y}{\theta_L^y}$ by industry y are presented in figure 4. If $\frac{\theta_H^y}{\theta_L^y} < 1$, non grdauates are more productive than graduates, in thta year and industry. $\frac{\theta_H^y}{\theta_L^y} > 1$, the opposite is true. Estimates show evidence of non linearities in the ratio's evolution: the ratio is well above one in 1987-1989 for most industries except services and transport and communications. From 1991-1994 to 2007-2009, high school graduates and non-graduates are about as equally

[3] Periods are 1987-1989, 1991-1993, 1994-1996, 1997-1999, 2000-2003, 2004-2006, 2007-2009, 2010-2012, 2013-2015, 2016-2017. Since data for years 1990 and 2001 are missing, the last time period spans only two years.

productive in all industries. Relative productivity of high school graduates increases dramatically in 2010-2013. This increase is heterogeneous in industries: services, manufacturing and transport and communications being the ones in which it increases most.

Figure 4: Estimated education productivities ratio

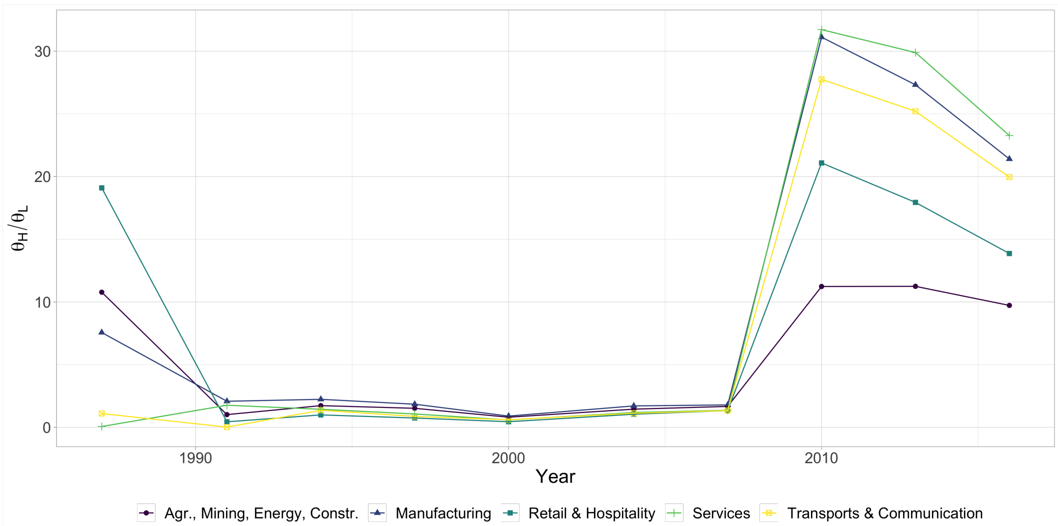


Figure 5 shows the evolution of age productivities $\lambda_{H,a}$ and $\lambda_{L,a}$ by age group and over time. Because $\lambda_{H,a}$ and $\lambda_{L,a}$ sum to one over age groups in any given year, they can be interpreted as shares of each age group in total labor input by high school graduates and non graduates. Estimates $\lambda_{L,a}$ are fairly stable over the period, and middle-workers make up most of the labor input for non high school graduates. Estimates $\lambda_{H,y}$ and $\lambda_{H,m}$ increase steadily until the early 2000s, but high school graduates senior workers input remains the most productive of the three.

Figure 5: Estimated age productivities

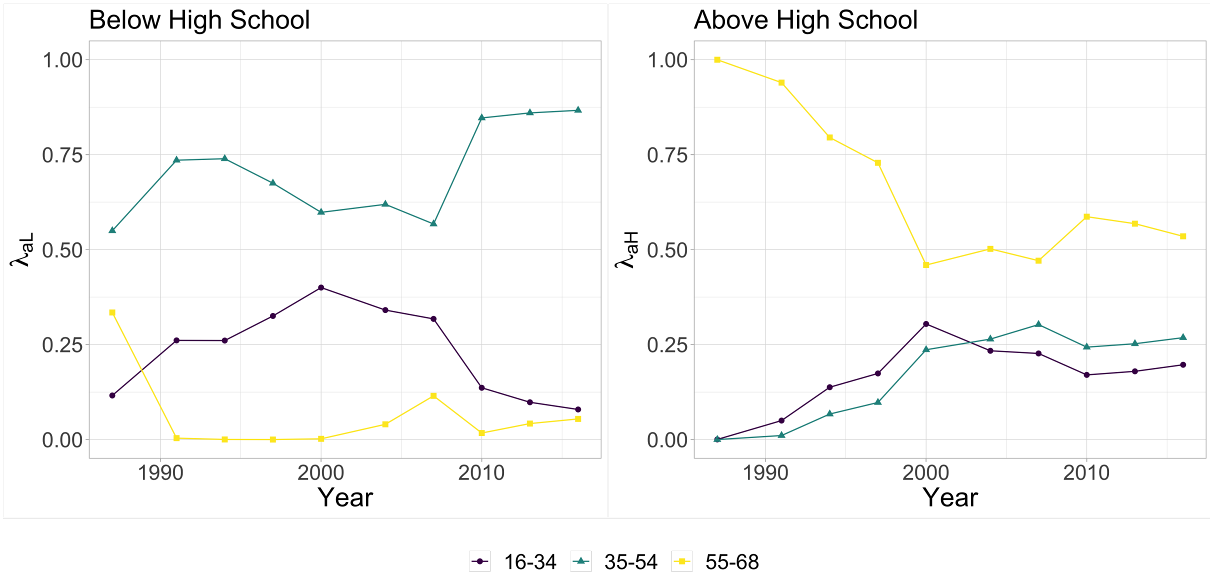
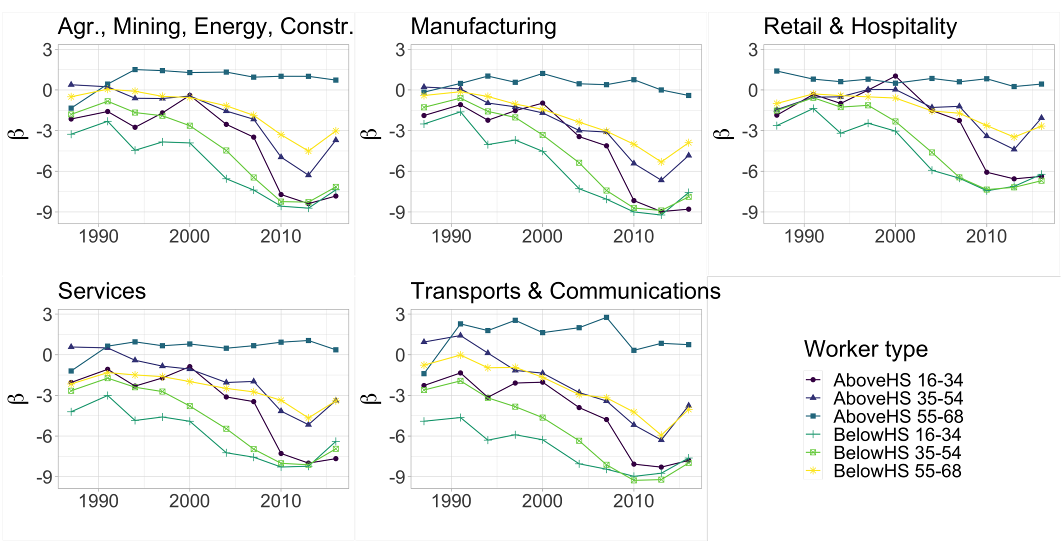


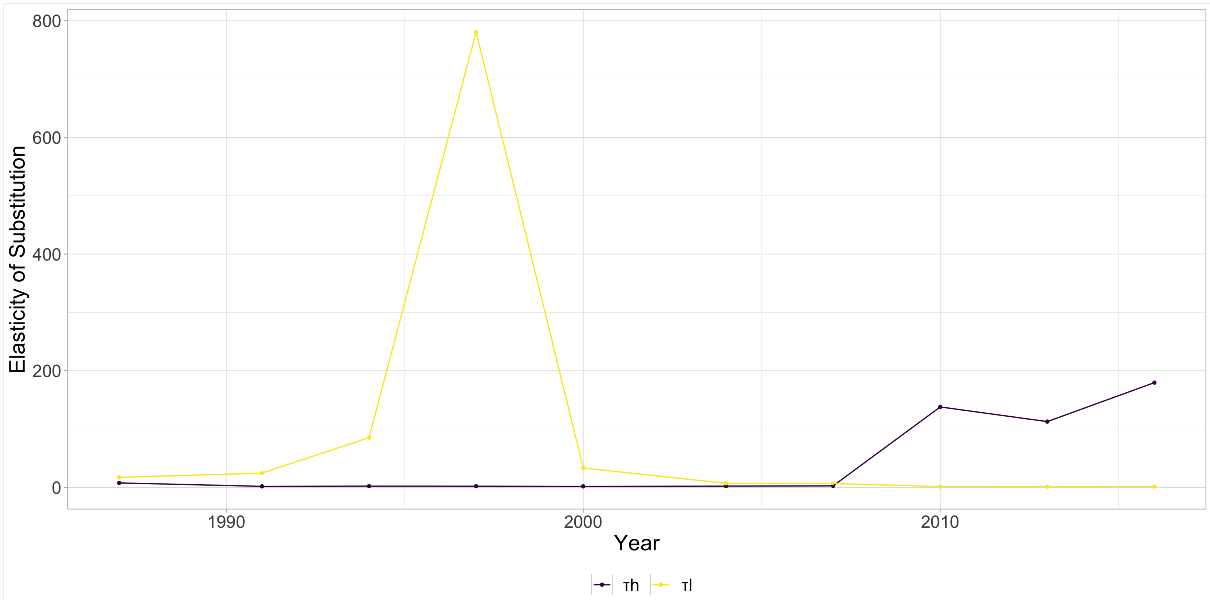
Figure 6 presents the change in worker preferences for firms β_x^y in euros per hours worked. All education levels and age groups display a disutility for working in all industries, except senior high school graduates. All groups experience negative change in their preferences, with varying intensity: young graduates being the ones who experience the steepest decline in most industries.

Figure 6: Estimated worker preferences



Finally, figure 7 presents estimated elasticities of substitution between ages. Both τ_H and τ_L are above one, meaning age groups are gross substitutes in both education levels. Substitution between age groups is generally higher among non high school graduates than graduates, although this is inverted in the last period 2016-2017.

Figure 7: Estimated elasticities of substitution



I find estimates for σ, the elasticity of substitution between education levels, between 425 in 1987 and 161 in 2016. In other words, high school graduates and non graduates are almost perfectly substitutable in Portugal over the period. This finding makes sense with regards to my estimation method, which unlike most of the literature estimates accounts for all firms in the economy, including those who employ only workers of a certain education level. These firms make up 78.5% and 63.1% of the total number of firms in the sample, in 1987 and 2016 respectively.

Summary. Structural estimates suggest that changes in relative productivity and in worker preferences may both account for the rise in sorting between young high school graduates and the service and transport and communications industry. The increase in relative productivity in the sectors has driven up demand for high school graduates, while the increase in substitutability between age groups have made it more profitable for firms to hire young graduates, who are increasingly more numerous on the labor market compared to their more senior peers. Meanwhile, young high school graduates preference for transports and communications has remained high relative to other age groups and education level, fuelling the increased sorting. The same observation does not apply to the service industry, for which young high school graduates' dislike has increased over time compared to other age groups.

5.2 Model Predictions

Table 1 compares the slopes of observed and predicted sorting over time. Slopes are obtained by fitting a time trend to the log of relative education supply in each age group and industry. They can be interpreted as average increase in sorting strength (measured as change in relative supply within an industry) over the period: for instance relative supply increases by on average 127.4% every period in the 16-34 age groups and the primary industries. Model predictions fit the manage to fit the changes in the data relatively well.

Table 1: Average yearly change in log relative supply of graduates versus non graduates - Data versus model predictions

	16-34		35-54		55-68	
Industry	Data	Prediction	Data	Prediction	Data	Prediction
Agr., Mining, Energy, Constr.	1.274	0.602	0.976	0.626	1.178	1.012
Manufacturing	1.478	1.842	1.222	1.94	0.928	1.525
Retail, Hospitality	1.16	1.207	1.315	1.364	0.933	1.133
Transports, Communication	1.391	2.066	1.748	2.46	1.454	2.333
Services	1.03	1.603	1.05	1.758	0.625	1.358
Overall	0.279	0.204	0.226	0.192	0.181	0.209

Table 2 compares the slopes of observed and predicted wage premium over time. Slopes are obtained by fitting a time trend to the log of wage gaps between education levels in each age group and industry. The model manages to qualitatively match the decrease in high school wage premium.

Table 2: Average yearly change in log wage premium of graduates versus non graduates - Data versus model predictions

	16-34		35-54		55-68	
Industry	Data	Prediction	Data	Prediction	Data	Prediction
Agr., Mining, Energy, Constr.	1.274	0.602	0.976	0.626	1.178	1.012
Manufacturing	1.478	1.842	1.222	1.94	0.928	1.525
Retail, Hospitality	1.16	1.207	1.315	1.364	0.933	1.133
Transports, Communication	1.391	2.066	1.748	2.46	1.454	2.333
Services	1.03	1.603	1.05	1.758	0.625	1.358
Overall	-0.035	-0.014	-0.018	0.024	-0.018	0.097

5.3 Counterfactuals

There are four categories of inputs that determine optimal matching and wage and that change over time: the number of workers of each type, the number of firms in each sector, production function parameters and worker preferences parameters. The first two are observed directly in the data and the last two are estimated. In the counterfactuals exercises that follow, I vary each one of the four inputs, holding all other three fixed between 1987-1989 and 2016-2017. The first counterfactual keeps the shares of each sector, production parameters and worker preferences constant to their 1987-1989 levels but lets the worker demography, both in terms of age group and education level, vary as it has in the data between 1987-1989 and 2016-2017. The second counterfactuals holds production parameters, worker preferences and worker demography fixed but lets sector shares vary. The third and fourth counterfactuals vary only production parameters and worker preferences, respectively.

The two object of interests are education-sector sorting and high school wage premium. The model makes predictions on both of these through equilibrium μ and w. Sorting between education and sector is defined as the log ratio of high school graduates over non-graduates employed in a sector y, for a given age group a:

$$\log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma,\beta,n,m)}{\sum_k k_{L,a}\mu_{yk}(\Gamma,\beta,n,m)}$$

where μ is the predicted matching. Therefore the change in sorting between two periods t and s is computed as

$$\Delta S_{y,a}^{s,t} = \log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^s,\beta^s,n^s,m^s)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^s,\beta^s,n^s,m^s)} - \log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^t,\beta^t,n^t,m^t)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^t,\beta^t,n^t,m^t)}$$

Change $\Delta S_{y,a}^{s,t}$ can be decomposed into the following:

$$\begin{aligned}
\Delta S_{y,a}^{s,t} =& \underbrace{\log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^s,\beta^s,n^s,m^s)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^s,\beta^s,n^s,m^s)} - \log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^t,\beta^t,n^s,m^t)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^t,\beta^t,n^s,m^t)}}_{\text{Change due to }\Gamma,\beta,m} \\
&+ \underbrace{\log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^t,\beta^t,n^s,m^t)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^t,\beta^t,n^s,m^t)} - \log \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^t,\beta^t,n^t,m^t)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^t,\beta^t,n^t,m^t)}}_{\text{Change due to }n} \qquad (15) \\
=& \Delta S_{y,a}^{s,t}(\Gamma,\beta,m) + \Delta S_{y,a}^{s,t}(n)
\end{aligned}$$

Decomposition (15) shows how the contribution of each input to changes sorting by sector and age bin can be isolated. Hence each of the four counterfactual exercises can be used to evaluate the change in sorting due to a single input $\Delta S_{y,a}^{s,t}(n)$, $\Delta S_{y,a}^{s,t}(m)$, $\Delta S_{y,a}^{s,t}(\Gamma)$ and $\Delta S_{y,a}^{s,t}(\beta)$, versus the change due to the rest of the inputs together, with t = 1987-1989 and s = 2016-2017.

Similarly, wage premium in a given sector y and age group a is

$$\log\left(\frac{\sum_k k_{H,a}\mu_{yk}(\Gamma,\beta,n,m)w_{\{H,a\}yk}(\Gamma,\beta,n,m)}{\sum_k k_{H,a}\mu_{yk}}\right) - \log\left(\frac{\sum_k k_{L,a}\mu_{yk}(\Gamma,\beta,n,m)w_{\{L,a\}yk}(\Gamma,\beta,n,m)}{\sum_k k_{L,a}\mu_{yk}}\right)$$

And the same decomposition as (15) can be performed, keeping in mind that changes in the wage premium can come both from changes in the wage schedule and variations in matching composition. Total change in average wage $\Delta W^{s,t}_{y,a}$ by age group and sector between periods t and s is:

$$\Delta W^{s,t}_{y,a} = \Delta W^{s,t}_{y,a}(\Gamma,\beta,m) + \Delta W^{s,t}_{y,a}(n) \tag{16}$$

The three tables below present counterfactual exercises for sorting in each age group and industry. The 1987-2017 change in sorting is computed as the predicted ratio of relative education supply in 2016-2017 over ratio of relative education supply in 1987-1989. I then run four counterfactuals, in which all of channels are fixed to their 1987-1989 levels, except one. These channels are relative education supply on the labor market, industry shares in the total number of firms, production parameter $(\theta,\lambda,\sigma,\tau)$, and worker preferences β. I compute changes in sorting in each of these four counterfactuals exercises in the same way as in the baseline and present the results in the tables below. They clearly suggest changes in education supply are driving the change in sorting, while changes in industry shares has a mitigating effect on sorting changes. Production parameters also have a small positive effect, while worker preferences effect is small and ambiguous depending on wage group.

Table 3: Education supply, industry composition, production parameters, and worker preferences contribution to changes in sorting, 16-34 age group

Industry	1987-2017 change	Education supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	1.301	2.302	-0.138	-0.695	-0.429
Manufacturing	3.198	2.432	-0.13	-0.586	-0.116
Retail, Hospitality	2.242	2.118	-0.172	-0.48	-0.362
Transports, Communication	3.715	2.022	-0.037	-0.827	0.503
Services	2.757	1.955	-0.003	-0.774	-0.278
Overall	1.987	2.172	-0.026	-0.663	-0.518

Table 4: Education supply, industry composition, production parameters, and worker preferences contribution to changes in sorting, 35-54 age group

Industry	1987-2017 change	Education supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	1.288	2.39	-0.259	-0.437	-0.371
Manufacturing	3.217	2.366	-0.155	-0.451	-0.072
Retail, Hospitality	2.232	2.21	-0.224	-0.327	-0.457
Transports, Communication	4.207	2.412	-0.115	-0.732	1.581
Services	2.823	1.857	-0.001	-0.682	-0.177
Overall	1.978	2.15	-0.033	-0.486	-0.461

Table 5: Education supply, industry composition, production parameters, and worker preferences contribution to changes in sorting, 55-68 age group

Industry	1987-2017 change	Education supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	1.63	2.48	-0.185	0.277	0.165
Manufacturing	2.611	2.117	-0.087	0.087	-0.224
Retail, Hospitality	1.939	1.985	-0.424	0.179	-0.789
Transports, Communication	4.332	1.811	-0.469	-0.202	4.026
Services	2.572	0.929	-0.065	-0.198	0.061
Overall	1.817	1.817	-0.105	0	-0.016

The three next tables present changes in wage premium between 1987-1989 and 2016-2017 for each counterfactual exercise, by industry and age group. Because the model is not restricted to strictly positive wages, it predicts in some counterfactuals negative wages, most notably when all inputs are fixed expect demographics for young and middle-aged workers. This points to the predominant role played by demographics: if neither industry composition, nor firm demand not worker preferences had changed, the education expansion would have driven high school graduates wages to negative levels. It is not the case for senior workers however, whose wage premium would still have increased in that scenario. Senior high school graduates would have suffered from changes in industry composition or changes in worker preferences alone however.

Table 6: Education supply, industry composition, production parameters, and worker preferences contribution to changes in wage premium, 16-34 age group

Industry	1987-2017 change	Education supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	-0.812	*	-0.703	**	-0.257
Manufacturing	0.001	*	-0.217	0.82	0.284
Retail, Hospitality	-1.15	*	-1.047	**	-0.371
Transports, Communication	0.197	*	-0.183	0.585	0.779
Services	-0.078	*	-0.334	1.855	-0.053
Services	-0.497	*	-0.589	2.485	-0.021

*: Average wage for high school graduates is predicted to be negative

**: Average wage for non high school graduates is predicted to be negative

Table 7: Education supply, industry composition, production parameters, and worker preferences contribution to changes in wage premium, 35-54 age group

Industry	1987-2017 change	Education supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	0.221	*	-0.884	**	0.156
Manufacturing	0.616	-6.16	-0.359	0.845	0.354
Retail, Hospitality	-0.625	-4.891	-0.707	**	-0.483
Transports, Communication	0.638	*	-0.316	0.631	1.01
Services	0.392	*	-0.327	1.396	0.102
Services	0.025	*	-0.536	1.408	-0.023

*: Average wage for high school graduates is predicted to be negative

**: Average wage for non high school graduates is predicted to be negative

Table 8: Education supply, industry composition, production parameters, and worker preferences contribution to changes in wage premium, 55-68 age group

Industry	1987-2017 change	Education supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	0.227	2.375	*	0.149	*
Manufacturing	1.014	2.336	*	0.544	*
Retail, Hospitality	-0.137	1.969	*	-0.372	*
Transports, Communication	0.944	2.49	*	0.365	-0.702
Services	0.836	2.325	*	0.369	*
Services	0.405	2.161	*	0.077	*

*: Average wage for high school graduates is predicted to be negative

**: Average wage for non high school graduates is predicted to be negative

6 Comparison to Card and Lemieux (2001)

Katz and Murphy (1992) and Card and Lemieux (2001) have shown that the CES production function parameters are identified from assuming that labor is optimally supplied to the economy and that wages are competitive, that is assuming that in each year t a representative firm solves

$$\max_{H_a, L_a} \gamma(t) - \sum_{a\in\{y,m,s\}} H_a w_{H,a} - \sum_{a\in\{y,m,s\}} L_a w_{L,a} \tag{17}$$

where $\gamma(t)$ is the CES production function described in section **??** with no dependence on firm type, as in this set up I assume a single representative firm. I also assume in this section that elasticities of substitution τ^H, τ^L, σ, as well as age productivity parameters $(\lambda_{H,a})_a$, $(\lambda_{L,a})_a$ do not vary with time. Wages are competitive and equal to marginal productivity:

$$\begin{aligned} w_{H,a}(t) &= \lambda_{H,a}^{\frac{\tau_H - 1}{\tau_H}} H_a(t)^{-\frac{1}{\tau^H}} \times \theta_H(t)^{\frac{\sigma-1}{\sigma}} H(t)^{\frac{1}{\tau_H} - \frac{1}{\sigma}} \times \gamma(t)^{\frac{1}{\sigma}} \quad \forall a \in \{y, m, s\}, \\ w_{L,a}(t) &= \lambda_{L,a}^{\frac{\tau_L - 1}{\tau_L}} L_a(t)^{-\frac{1}{\tau^L}} \times \theta_L(t)^{\frac{\sigma-1}{\sigma}} L(t)^{\frac{1}{\tau_L} - \frac{1}{\sigma}} \times \gamma(t)^{\frac{1}{\sigma}} \quad \forall a \in \{y, m, s\}. \end{aligned} \tag{18}$$

Which results in relative wage equations:

$$\begin{aligned} \log\left(\frac{w_{H,a}(t)}{w_{H,a'}(t)}\right) &= \frac{\tau_H - 1}{\tau_H} \log\left(\frac{\lambda_{H,a}}{\lambda_{H,a'}}\right) - \frac{1}{\tau^H} \log\left(\frac{H_a(t)}{H_{a'}(t)}\right), \\ \log\left(\frac{w_{L,a}(t)}{w_{L,a'}(t)}\right) &= \frac{\tau_L - 1}{\tau_L} \log\left(\frac{\lambda_{L,a}}{\lambda_{L,a'}}\right) - \frac{1}{\tau^L} \log\left(\frac{L_a(t)}{L_{a'}(t)}\right). \end{aligned} \tag{19}$$

Restricting $(\lambda_{H,a})_a$, $(\lambda_{L,a})_a$'s variation in time, and adding a stochastic shock to account for measurement errors in observed wage and hours worked, relative age productivity and age elasticities of substitution can therefore be estimated by ordinary least squares through equations:

$$\begin{aligned} \log\left(\frac{w_{H,a}(t)}{w_{H,a_0}(t)}\right) &= d_{H,a,a_0} - \frac{1}{\tau^H} \log\left(\frac{H_a(t)}{H_{a_0}(t)}\right) + u_{H,a,a_0} \\ \log\left(\frac{w_{L,a}(t)}{w_{L,a_0}(t)}\right) &= d_{L,a,a_0} - \frac{1}{\tau^L} \log\left(\frac{L_a(t)}{L_{a_0}(t)}\right) + u_{L,a,a_0} \end{aligned} \tag{20}$$

where a_0 is the reference age category. Age productivities $\lambda_{H,a}$, $\lambda_{L,a}$ can then be retrieved from fixed effect d_{H,a,a_0}, d_{L,a,a_0} using normalization conditions (10).

Estimates for aggregate labor inputs $H(t)$ and $L(t)$ can be computed from estimated age productivities and elasticities of substitution. First order conditions (18) also give an expression for relative wage across education levels:

$$\log\left(\frac{w_{H,a}(t)}{w_{L,a}(t)}\right) - \log\left(\frac{\left(\lambda_{H,a}^{\tau_H-1}\frac{H(t)}{H_a(t)}\right)^{\frac{1}{\tau^H}}}{\left(\lambda_{L,a}^{\tau_L-1}\frac{L(t)}{L_a(t)}\right)^{\frac{1}{\tau^L}}}\right) = \frac{\sigma-1}{\sigma}\log\left(\frac{\theta_H(t)}{\theta_L(t)}\right) - \frac{1}{\sigma}\log\left(\frac{H(t)}{L(t)}\right). \quad (21)$$

Assume $\log\left(\frac{\theta_H(t)}{\theta_L(t)}\right)$ follows a linear time trend. Plugging in previously estimated age productivities and elasticities of substitution and adding measurement error gives us equation

$$\log\left(\frac{w_{H,a}(t)}{w_{L,a}(t)}\right) - \hat{f} = l(t) - \frac{1}{\sigma}\log\left(\frac{H(t)}{L(t)}\right) + v_{a,t} \quad (22)$$

where $l(t)$ is a linear function of time and $\hat{f}$ is estimated from equations (20).

Weighted Least Square estimation of equations (20) and (22) are presented in table 9 and 10. The weights used are the inverse sampling variance of estimated wage gaps[4]. Labor input from any given education level and age bin is computed as the total sum of hours workers per month in a year. Average wage premia between age and within education are used as outcome variable in equation (20) and computed yearly and by education level by regressing individual wages on a dummy for age, plus fixed effects for gender, industry and region, to control for composition effects. Average wage premia between education levels and within ages are computed in the same fashion.

Table 9: Estimated age productivities and elasticities of substitution - Reduced Form

	Below High School	Above High School
τ	15.907	15.301
	(2.216)	(2.427)
λ_y	0.332	0.331
	(0)	(0.001)
λ_m	0.333	0.332
	(0)	(0)
λ_s	0.335	0.338
	(0)	(0.001)
R^2	0.994	0.972
Obs.	58	58

Estimated age elasticities of substitution τ in Portugal from 1987 to 2017 are higher than

[4] In equation (22), I weight by the inverse of the sum of the wage gaps and $\hat{f}$ inverse sampling variance

estimates found by Card and Lemieux (2001) for the US, the UK and Canada from the 1970s to the early 1990s, which are between 4 and 6. This reflects the lesser impact of movements in relative age group supply on age group wage differential in Portugal than in the US, UK and Canada. Estimated age productivities are very similar between education levels. They are also balanced between age groups, which suggests no age group is much more productive than another.

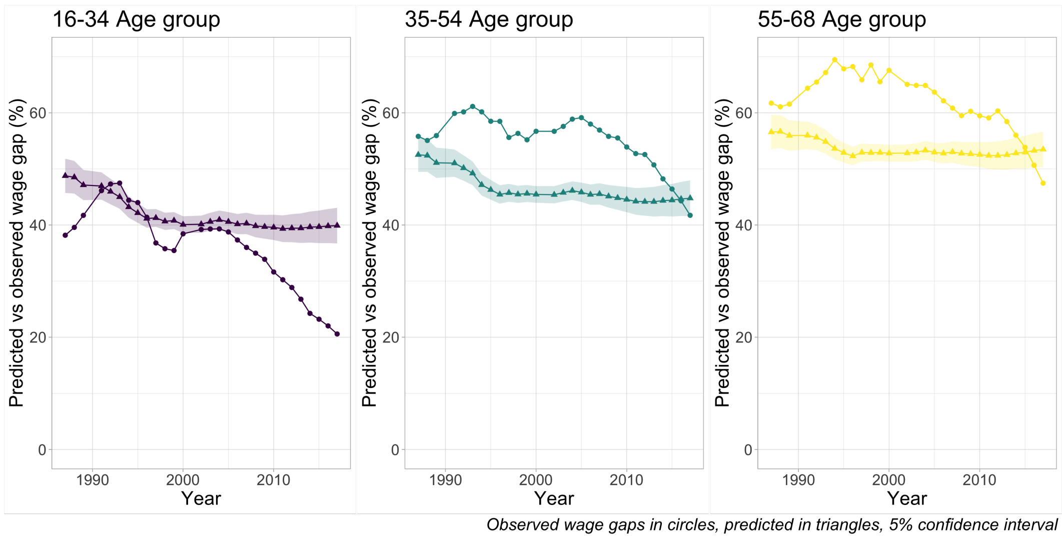
Table 10: Estimated education productivity growth and elasticity of substitution - Reduced Form

σ	4.933
	(0.151)
$\log \frac{\theta_H}{\theta_L}$	0.018
	(0.001)
R^2	0.974
Obs.	87

Elasticity of substitution between workers below and above high school is also higher in Portugal than what is found by Katz and Murphy (1992) for the US and Card and Lemieux (2001) for the UK and the US, who has estimates between 2 and 2.5. However Card and Lemieux (2001) find no significant effect of relative labor supply on relative wage between education levels in Canada, suggesting a very high substitutability of graduates and non graduates in that country. their analysis also focuses on college versus high school graduates, which is not directly comparable to my analysis on high school graduates and non graduates, who appear to be more substitutable than college graduates and non graduates. Like Katz and Murphy (1992) and Card and Lemieux (2001), I find evidence of skill-biased technological change in Portugal over the period, as relative productivity between education groups increases by 1.6% every year. This is in the range of what Card and Lemieux (2001) find for the US, UK and Canada.

This analysis informs on the large substitutability of workers between age groups and education levels, as well as the slow but significant high school biased technological change occurring in the Portuguese economy between 1987 and 2017, under simple assumptions on supply and demand. Its conclusion is that it is the increase in relative supply of high school graduates that causes the decrease in wage premium, in particular for young workers, who experience a more important rise in relative supply. 8 presents the predicted wage gaps by age group, against observed wage gaps in the data. Coefficients estimated with Card and Lemieux (2001)'s method match fairly well the wage premium at the beginning and end of the period, except for young workers. However predicted wage is flatter over the period than it is the data, suggesting that imposing linearity in the evolution of relative education productivity, as well as constant age productivity across time, restricts the estimation too much.

Figure 8: Predicted wage gaps between high school graduates and non-graduates of same age



7 Conclusion

To jointly explain changes in sorting between workers and firms, and the decreasing wage premium on the Portuguese labor market, I build a static model of one-to-many matching with transferable utility. Using predictions for both wages and joint distribution of firms and workforces, I am able to separately estimate worker preferences for firms and parameters for firms' nested CES production functions. Estimates suggest changes in sorting are driven by heterogeneity in sectors' relative demand over time, as well as changes in workers' preferences. They also suggest the decreasing high school wage premium is driven mainly by an increase in the relative supply of high school graduates to non-graduates.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis
1999. High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.

Almeida, A., H. Figueiredo, J. Cerejeira, M. Portela, C. Sa, and P. Teixeira
2017. Returns to Postgraduate Education in Portugal:Holding on to a Higher Ground? NIPE Working Paper 08/2017, NIPE - Universidade do Minho.

Autor, D., C. Goldin, and L. F. Katz
2020. Extending the Race between Education and Technology. *AEA Papers and Proceedings*, 110:347–351.

Autor, D. H., L. F. Katz, and A. B. Krueger
1998. Computing Inequality: Have Computers Changed the Labor Market?*. *The Quarterly Journal of Economics*, 113(4):1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane
2003. The Skill Content of Recent Technological Change: An Empirical Exploration*. *The Quarterly Journal of Economics*, 118(4):1279–1333.

Azevedo, E. M. and J. W. Hatfield
2018. Existence of Equilibrium in Large Matching Markets With Complementarities. SSRN Scholarly Paper ID 3268884, Social Science Research Network, Rochester, NY.

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(S1):S199–S247. Publisher: The University of Chicago Press.

Becker, G. S.
1973. A Theory of Marriage: Part I. *Journal of Political Economy*, 81(4):813–846.

Bikhchandani, S. and J. M. Ostroy
2002. The Package Assignment Model. *Journal of Economic Theory*, 107(2):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.
2021. Teams: Heterogeneity, Sorting, and Complementarity. *Working Paper*.

Bonhomme, S., T. Lamadon, and E. Manresa
2019. A Distributional Framework for Matched Employer Employee Data. P. 53.

Burdett, K. and D. T. Mortensen
1998. Wage Differentials, Employer Size, and Unemployment. *International Economic Review*, 39(2):257–273. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

Cahuc, P., F. Postel-Vinay, and J.-M. Robin
2006. Wage Bargaining with On-the-Job Search: Theory and Evidence. *Econometrica*, 74(2):323–364.

Card, D. and J. E. DiNardo
2002. Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. Working Paper 8769, National Bureau of Economic Research. Series: Working Paper Series.

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(2):705–746.

Cardoso, A. R.
2004. Jobs for Young University Graduates: Is It Worth Having a Degree?

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

Che, Y.-K., J. Kim, and F. Kojima
2019. Stable Matching in Large Economies. *Econometrica*, 87(1):65–110. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13547.

Choné, P. and F. Kramarz
2021. Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling. *CREST Working Paper*, P. 80.

Choo, E. and A. Siow
2006. Who Marries Whom and Why. *Journal of Political Economy*, 114(1):175–201.

Dupuy, A. and A. Galichon
2014. Personality Traits and the Marriage Market. *Journal of Political Economy*, 122(6):1271–1319.

Dupuy, A. and A. Galichon
2017. A Note on the Estimation of Job Amenities and Labor Productivity. SSRN Scholarly Paper ID 3004406, Social Science Research Network, Rochester, NY.

Fox, J., C. Yang, and D. H. Hsu
2018. Unobserved Heterogeneity in Matching Games. *Journal of Political Economy*, Vol 126, No 4.

Fox, J. T.
2009. Identification in Matching Games. SSRN Scholarly Paper ID 1422970, Social Science Research Network, Rochester, NY.

Galichon, A.
2016. *Optimal Transport Methods in Economics.* Princeton University Press.

Galichon, A. and B. Salanie
2020. Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. *SSRN Electronic Journal*.

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

Goodfellow, I., Y. Bengio, and A. Courville
2016. *Deep Learning.* MIT Press.

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

Hershbein, B. and L. B. Kahn
2018. Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. *American Economic Review*, 108(7):1737–1772.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow
2019. The Allocation of Talent and U.S. Economic Growth. *Econometrica*, 87(5):1439–1474. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA11427.

Katz, L. F. and K. M. Murphy
1992. Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Kelso, A. S. and V. P. Crawford
1982. Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica*, 50(6):1483–1504.

Kingma, D. P. and J. Ba
2017. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. arXiv: 1412.6980.

Lindenlaub, I.
2017. Sorting Multidimensional Types: Theory and Application. *The Review of Economic Studies*, 84(2):718–789.

McFadden, D.
1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics*.

Postel-Vinay, F. and J. Lise
2015. Multidimensional Skills, Sorting, and Human Capital Accumulation. 2015 Meeting Paper 386, Society for Economic Dynamics.

Sattinger, M.
1979. Differential Rents and the Distribution of Earnings. *Oxford Economic Papers*, 31(1):60–71. Publisher: Oxford University Press.

Sattinger, M.
1993. Assignment Models of the Distribution of Earnings. *Journal of Economic Literature*, 31(2):831–880. Publisher: American Economic Association.

Shimer, R. and L. Smith
2000. Assortative Matching and Search. *Econometrica*, 68(2):343–369.

Valletta, R. G.
2016. Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both? Working Paper 22935, National Bureau of Economic Research. Series: Working Paper Series.

A Data

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 1987 to 2017, based on firms declarations on their characteristics and their employees'. Both workers and firms are identified across time by a unique identifier.

I use information on firm industry, worker's age and education level Industries are provided as "economic activity", up to 3 digit level. Because of classification changes at the 2 and 3 digits level over time, I use the one digit level classification, to keep consistency over the years. I exclude firms whose economic activity at the 1 digit level are unknown. Worker education is provided as a 3 digits classification, out of which I aggregate 9 levels: no schooling, primary schooling 1 (up to 10 years old), primary schooling 2 (up to 13 years old), primary schooling 3 (up to 15 years old), completed high school, some higher education, bachelor, masters and PhD. Worker age is used directly without further cleaning. I exclude from the sample any worker whose education level of age is unknown (3.9% of observations per year on average)

I also use information on wages and number of hours worked. Wage is provided as a average monthly earnings, that accounts 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 he year). I consider the sum of base and extra hours as my measure for number of hours worked per month. I divide monthly wage by monthly hours to obtain a measure of hourly wage, and deflate it. Real hourly wage is my final measure of wage. I exclude from the sample any worker who has worked zero hours or earned zero wage over the year (11.5% of observations per year on average). These are mainly, in my understanding, workers on sick leave, maternity leave, or sabbatical that do not contribute to firm production in that year. I also exclude from the sample any workers who are strictly under 16 or above 68 (the retirement age in Portugal)

Additionally, I exclude any observation with a missing or 0 worker ID (3.5% of observations per year on average). I am also faced with an issue of duplicate worker IDs which, even though it is minor in the sample later years (about 4.8% of observations per year on average from 2007 to 2017, including 0 IDs), it is much more serious in the earlier years (about 19% of the sample in 1987, including 0 IDs). I suspect these to be encoding mistakes that relate to actual different workers. Some can also be workers who hold two different jobs (for instance an employee somewhere who also have a self-employed activity). Because I do not use the panel aspect of the data, and therefore encoding mistakes in workers ID are not a problem in my analysis, I keep most duplicates, only removing observations who appear more than 5 times in any given year (an average 6.1% of observations per year, less than 1% of the dataset starting in 2007). I also exclude from the sample any worker who is self-unemployed, in unpaid family care, or labelled under "other" employment contract (7.1% of observations per year on average). 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 on their own represent about 1% of the dataset.

B Proofs

Proposition 1

Proof. Let $Z_1 = \max_{y,k}\{u_{xyk} + \xi\epsilon_{iyk}\}$ and $Z_2 = \max_k\{v_{yk} + \xi\eta_{jk}\}$. The proof consists is showing that Z_1 follows a Gumbel distribution with expectation $\xi \log \sum_y \sum_k \exp\left(\frac{u_{xyk}}{\xi}\right)$ and Z_2 follows a Gumbel distribution with expectation $\xi \log \sum_k \exp\left(\frac{v_{yk}}{\xi}\right)$.

$$\begin{aligned}
\mathbb{P}[Z_1 \leq c] =& \mathbb{P}\left[\epsilon_{iyk} \leq \frac{c - u_{xyk}}{\xi}\ \forall y, k\right] \\
=& \prod_{y,k} \mathbb{P}\left[\epsilon_{iyk} \leq \frac{c - u_{xyk}}{\xi}\right] \\
=& \prod_{y,k} \exp\left(-\exp\left(\frac{u_{xyk} - c}{\xi}\right)\right) \\
\Rightarrow \log \mathbb{P}[Z_1 \leq c] =& -\sum_{y,k} \exp\left(\frac{u_{xyk} - c}{\xi}\right) \\
=& -\exp\left(\frac{-c + \log \sum_{y,k} \exp(u_{xyk})}{\xi}\right)
\end{aligned}$$

And a similar reasoning shows:

$$\mathbb{P}[Z_2 \leq c] = -\exp\left(\frac{-c + \log \sum_k \exp(v_y(k))}{\xi}\right)$$

Hence up to the Euler-Mascheroni constant, Z_1 follows a Gumbel distribution with expectation $\xi \log \sum_y \sum_k \exp\left(\frac{u_{xyk}}{\xi}\right)$ and Z_2 follows a Gumbel distribution with expectation $\xi \log \sum_k \exp\left(\frac{v_{yk}}{\xi}\right)$. □

Proposition 2

Proof. Following McFadden (1974), Choo and Siow (2006), the probability that worker x

chooses option $\bar{y}, \bar{k}$ is

$$\begin{aligned}\mathbb{P}\left[\bar{y}, \bar{k} = \arg\max u_{xyk} + \xi\epsilon_{yk}\right] &= \mathbb{P}\left[\xi\epsilon_{yk} \leq u_{x\bar{y},\bar{k}} - u_{xyk} + \xi\epsilon_{\bar{y}\bar{k}}\, \forall y, k\right] \\ &= \int \prod_{y,k} \exp\left(-\exp\left(\frac{u_{x\bar{y},\bar{k}} - u_{xyk} + \epsilon}{\xi}\right)\right) \exp(-\epsilon) \exp\left(-\exp(-\epsilon)\right)\, d\epsilon \\ &= \frac{\exp\left(\frac{u_{xyk}}{\xi}\right)}{1 + \sum_{y,k} \exp(\frac{u_{xyk}}{\xi})}\end{aligned}$$

A similar derivation applied on the firm side. □

Theorem 1 Based on Gretsky et al. (1992) and Galichon and Salanie (2020).

Proof. Consider the following problem over the sum of worker welfare $\int_i u_i\, di$ and firm welfare $\int_j v_j\, dj$:

$$\begin{aligned}&\inf_{u,v} \int_i u_i\, di + \int_j v_j\, dj \\ &\text{s.t } \sum_x \sum_{i:x_i=x}^{k_x} u_i + v_j \geq \Phi_{y_jk} + \xi \sum_x \sum_{i:x_i=x}^{k_x} \epsilon_{iy_jk} + \xi\eta_{jk} \quad \forall k, j \\ &\quad u_i \geq \xi\epsilon_{i0}\end{aligned} \tag{23}$$

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

$$\begin{cases} u_i = \max_{y,k}\{u_{x_iyk} + \xi\epsilon_{iyk}\} \\ v_j = \max_k\{v_{y_jk} + \xi\eta_{jk}\} \end{cases}$$

Then (u, v) satisfies (23)'s constraints.
Reciprocally, fix any u_i, v_j that satisfy the constraints in this problem and define
Let

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

Then the constraint in problem (23) becomes $\sum_x k_x u_{xyk} + v_{yk} \geq \Phi_{yk}$.

Applying the law of large numbers, we get that (23) is equivalent to

$$\begin{aligned}&\min_{u,v} \sum_x n_x G_x(u_x) + \sum_y m_y H_y(v_y) \\ &\text{s.t } \sum_x k_x u_{xyk} + v_{yk} = \Phi_{yk} \quad \forall k, y \\ &\quad u_{x0} = 0\end{aligned} \tag{24}$$

By complementary slackness condition, solving problem (23) with $u_{xyk} = \alpha_{xyk} + w_{xyk}$ and $v_{yk} = \gamma_{yk} - \sum_x k_x w_{xyk}$ yields equilibrium wage. Equilibrium supply and demand $S^x_{yk} = k_x D^y_k$ obtain as the Lagrange multiplier μ_{yk} on constraint $\sum_x k_x u_{xyk} + v_{yk} \geq \Phi_{yk}$.

□

Proof. **Theorem 2**
Rewrite problem (4) as saddle-point:

$$\begin{aligned}
&\min_{u,v} \max_{\mu} \sum_x n_x G_x(u_x) + \sum_y m_y H_y(v_y) \\
&\quad + \sum_{y,k} \mu_{yk} \left(\Phi_{yk} - \sum_x k_x u_{xyk} - v_{yk} \right) + \sum_x S^x_0(-u_{x0}) \\
&= \max_{\mu} \sum_{y,k} \mu_{yk} \Phi_{ky} \\
&\quad - \sum_x n_x \max_u \left\{ \sum_y \frac{k_x \mu_{yk}}{n_x} u_{xyk} + \frac{S^x_0}{n_x} u_{x0} - G_x(u) \right\} \\
&\quad - \sum_y m_y \max_v \left\{ \sum_y \frac{\mu_{yk}}{m_y} v_{yk} - H_y(v) \right\} \\
&= \max_{\mu} \sum_{y,k} \mu_{yk} \Phi_{ky}\, dk \\
&\quad - \sum_x \sum_{y,k} k_x \mu_{yk} \log \frac{k_x \mu_{yk}}{n_x} - \sum_x S^x_0 \log \frac{S^x_0}{n_x} - \sum_{y,k} \mu_{yk} \log \frac{\mu_{yk}}{m_y}
\end{aligned}$$

where the last line is obtained through solving for Fenchel-Legendre transforms of G and H:

$$\begin{aligned}
G^*_x(\mu) &= \max_u \left\{ \sum_{y,k} \frac{k_x \mu_{yk}}{n_x} u_{xyk} + \frac{S^x_0}{n_x} u_{x0} - G_x(u) \right\} \\
H^*_y(\mu) &= \max_v \left\{ \sum_{y,k} \frac{\mu_{yk}}{m_y} v_{yk} - H_y(v) \right\}.
\end{aligned}$$

For which first order conditions are

$$\frac{k_x \mu_y(k)}{n_x} = \frac{\exp(u_{xy}(k))}{\sum_y \int \exp(u_{xy}(k))\, dk} \text{ and } \frac{\mu_y(k)}{m_y} = \frac{\exp(v_y(k))}{\int \exp(v_y(k))\, dk}$$

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

Problem (5) attains its maximum because its objective function is continuous in μ and (5) maximizes the social planner function on a bounded and closed set.

$\square$

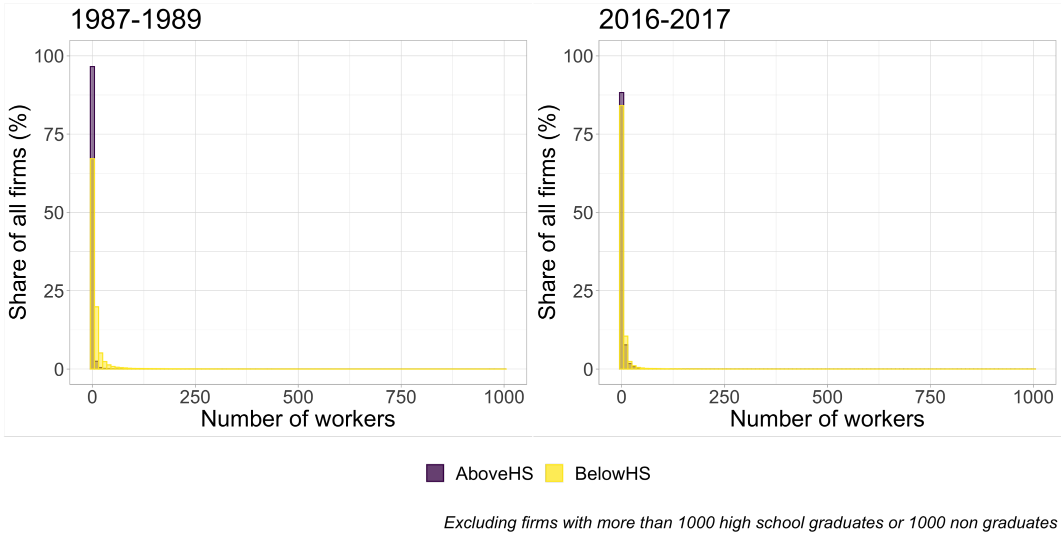
C Estimation

This section covers the details of log likelihood estimation. Subsection C.1 describes how to data is processed into observed matchings and wages by firm type and workforce, subsection C.2 presents the gradient descent algorithm used for log likelihood maximization and subsection C.3 provides details on log likelihood gradient computation.

C.1 Workforce Discretization

From *Quadros de Pessoal*, I build a workforce matched to each firm every year, by using firm identifiers provided in the data. I weigh workers by their number of monthly hours worked on average over ear, which is directly provided in the dataset. One full-time worker is equivalent to 174 hours worked per month (which is a 40 hours week). If for instance a worker has worked 180 hours per month, she counts as $\frac{180}{174} = 1.03$ full-time workers. The distribution of firms by number of high school graduates and non graduates employed is plotted in figure 9 in the periods 1987-1989 and 2016-2017. A firm is defined through a distinct firm identifier-year combination. 9 shows a large majority of firms are small firms. Many firms employ no high school graduates, especially at the start of the period: they represent 75.9% of firms in 1987-1989, and 37.0% of firms in 2016-2017. In contrast, firms who do not employ no high school graduates make up 2.6% and 26.4% of all firms, in 1987-1989 and 2016-2017 respectively. Firms who employ more than a thousand of workers at one education level are excluded from the graph, but not from the estimation. They represent 234 firms in 1987-1989 and 206 firms in 2016-2017.

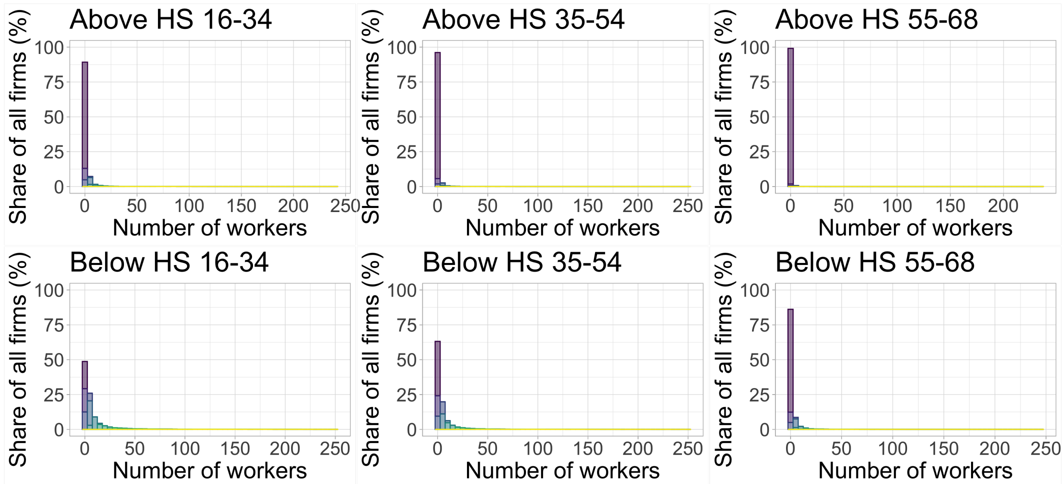
Figure 9: Firm distribution by number of high school graduates and non graduates employed



Performing the estimation requires to compute observed matching $\tilde{\mu}_{yk}$ and observed average wage $\tilde{W}_{xyk}$ by workforce k. The number of different observed workforces in the data is very large: there are 96612 combinations in 1987-1989 and 69314 in 2016-2017. Max likelihood computation requires to evaluate μ_{yk} and w_{xyk} on all observed workforces. To speed up the max likelihood computation, I cluster observed workforces into a smaller number of representative workforces. To do so, choose a number of bins B. For each worker type x, split the interval between 0 and k_x^{max} in B smaller intervals, where k_x^{max} is the largest observed number of type x workers employed by a firm. For each worker type x, the procedure yields B intervals, or clusters $[0, k_x^1), \ldots, [k_x^{b-1}, k_x^b), \ldots, [k_x^B, k_x^{max}]$. Each observed number of worker x employed by a firm falls into one of these intervals. I assign each observed number to a cluster. The representative number of workers for each cluster is $\frac{k_x^b - k_x^{b-1}}{2}$.

In the baseline estimation, $B = 15$. Intervals are split according to a logarithmic scale. The number of observed clusters is reduced to 10359 in 1987-1989 and 21871 in 2016-2017. As an illustration, figure 10 displays worker distribution across firms by type, and the clustering of workforce.

Figure 10: Worker type distribution and clusters, 1987-1989



Each color represents a different cluster.
Excluding firms employing more that 250 workers of each type.

C.2 Adam Algorithm

Adam is a first-order gradient-based optimization algorithm. It belong to the family of algorithms with adaptive learning rates. Their main benefit is speed: they use information given by the gradient to modify their learning rate, and hence improve convergence speed. In particular, Adam uses momentum, i.e. an exponentially moving average of past gradients, at each iteration. It also uses bias correction. Adam was first introduced by Kingma and Ba (2017). For a general presentation of the algorithm, see Goodfellow et al. (2016). The algorithm applied to the present problem goes as follows:

Set decay rates $\rho_1 = .9$, $\rho_2 = .999$, step $\epsilon = 1e - 2$, stabilizer $\delta = 1^{e-8}$ and tolerance $tol = 1^{e-4}$.

Initialize parameters to Γ_0, β_0

Initialize moment variables $s = 0$, $r = 0$ and time step $t = 0$

While $\max \left| \frac{\nabla_{\Gamma,\beta} l(\Gamma_t, \beta_t, n, m, s_t^2)}{l(\Gamma_t, \beta_t, n, m, s_t^2)} \right| > tol$

> **Compute** $s_t^2 = \frac{1}{W} \sum_x \sum_{y,k} \tilde{K}_{xyk} \left(\tilde{W}_{xyk} - w_{xyk}(\Gamma_t, \beta_t, n, m) \right)^2$
>
> **Compute** $g \leftarrow \nabla_{\Gamma,\beta} l(\Gamma_t, \beta_t, n, m, s_t^2)$
>
> **Update** $t \leftarrow t + 1$
>
> **Update** $s \leftarrow \rho_1 s + (1 - \rho_1) g$ and $r \leftarrow \rho_2 r + (1 - \rho_2) g \odot g$
>
> **Correct bias** in first moment $\hat{s} \leftarrow \frac{s}{1-\rho_1^t}$ and second moment $\hat{r} \leftarrow \frac{r}{1-\rho_2^t}$
>
> **Compute** update $\Delta(\Gamma, \beta) = \frac{\hat{s}}{\sqrt{\hat{r}} + \delta}$
>
> **Apply** update $(\Gamma_{t+1}, \beta_{t+1}) \leftarrow (\Gamma_t, \beta_t) + \epsilon \Delta(\Gamma, \beta)$

end While

C.3 Likelihood gradient

Applying Adam requires to compute likelihood gradient $\nabla_{\Gamma,\beta} l(\Gamma_t, \beta_t, n, m, s_t^2)$. Let $\omega \in (\Gamma, \beta)$ be any of the parameters governing firm production or workers' preferences. Log likelihood differential with respect to ω is

$$\begin{aligned}\frac{\partial l(\Gamma,\beta,n,m,s^2)}{\partial \omega} =& \sum_x \sum_{y,k} k_x \tilde{\mu}_{yk} \frac{\partial \log \mu_{yk}(\Gamma,\beta,n,m,s^2)}{\partial \omega} + \sum_x \tilde{S}_0^x \frac{\partial \log S_0^x(\Gamma,\beta,n,m,s^2)}{\partial \omega} \\ & - \sum_x \sum_{y,k} \tilde{K}_{xyk} \frac{\left(\tilde{W}_{xyk} - \frac{\partial w_{xyk}(\Gamma,\beta,n,m,s^2)}{\partial \omega}\right)^2}{2s^2}\end{aligned}$$

Where

$$\begin{aligned}\frac{\partial \log \mu_{yk}(\Gamma,\beta,n,m,s^2)}{\partial \omega} &= \frac{1}{1+\sum_x k_x}\left(\frac{\partial \Phi_{yk}}{\partial \omega} - \sum_x k_x \frac{\partial U_x}{\partial \omega} - \frac{\partial V_y}{\partial \omega}\right) \\ \frac{\partial \log S_0^x(\Gamma,\beta,n,m,s^2)}{\partial \omega} &= -\frac{\partial U_x}{\partial \omega} \\ \frac{\partial w_{xyk}(\Gamma,\beta,n,m,s^2)}{\partial \omega} &= \frac{1}{1+\sum_x k_x}\left(\frac{\partial \Phi_{yk}}{\partial \omega} - \sum_x k_x \frac{\partial U_x}{\partial \omega} - \frac{\partial V_y}{\partial \omega}\right) - \frac{\partial \alpha_{xyk}}{\partial \omega} + \frac{\partial U_x}{\partial \omega}\end{aligned}$$

$\frac{\partial \Phi_{yk}}{\partial \omega}$ and $\frac{\partial \alpha_{xyk}}{\partial \omega}$ can be computed directly given their assumed functional forms. $\frac{\partial U_x}{\partial \omega}$ and $\frac{\partial V_y}{\partial \omega}$ solve the following linear equations:

$$\begin{aligned}\sum_{y,k} \frac{k_x}{1+\sum_x k_x}\mu_{yk}\left(\sum_x k_x \frac{\partial U_x}{\partial \omega} + \frac{\partial V_y}{\partial \omega}\right) &= \sum_{y,k} \frac{k_x}{1+\sum_x k_x}\mu_{yk}\frac{\partial \Phi_{yk}}{\partial \omega} \quad \forall x \\ \sum_{k} \frac{1}{1+\sum_x k_x}\mu_{yk}\left(\sum_x k_x \frac{\partial U_x}{\partial \omega} + \frac{\partial V_y}{\partial \omega}\right) &= \sum_{k} \frac{1}{1+\sum_x k_x}\mu_{yk}\frac{\partial \Phi_{yk}}{\partial \omega} \quad \forall y\end{aligned}$$

Which are obtained by differentiating marginal conditions (8).