Education Expansion, Sorting, and the Decreasing Education Wage Premium

Pauline Corblet*

December 23, 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

Between the 1970s and today, many economies both in the developed and developing world have experienced an increase in their educated labor supply. As a result, the ratio of educated workers (whether high school or college-educated) to uneducated workers present on labor markets has risen. The shift in labor supply's education level has induced broad changes on labor markets, both in terms of workers' allocation to firms and wage structure. Specifically, firms in sectors where the need for educated workers is high profit from an increase in educated labor supply and hire a larger number of workers. But if demand for educated labor remains constant, the increase in the relative supply of educated workers pressures their own wage downwards. However worker-firm allocation and wage structure are also impacted by changes on the demand side, as sectors' demand for education evolves. This paper seeks to address the lack of a theoretical framework to understand these changes and proposes a novel model of matching on the labor market in which a single firm matches with several workers. The model is structurally estimated on Portuguese matched employer-employee data. In doing so, I am able to quantify the impact of supply and demand changes on worker-firm allocation and wage structure.

The mechanisms driving matching between workers and firms and the resulting wage distribution are two-sided. On the one hand employed workers with various education and experience interact within the firm to produce an output, whose level depends on a production function that is particular to each firm's sector. Given their production function, firms seek to hire a workforce, which is a mix of workers with different characteristics, to maximize their profit. On the other hand, workers have preferences for the tasks performed on the job, which vary from one sector to another. Worker preferences impact which sector they are willing to work in. Given distributions of education and experience in the worker population and sectorial composition among firms, firm production requirements and worker preferences result in a given level of sorting and wage gaps. Sorting is the result of worker-firm allocation: it is the ratio of educated to non-educated workers in each sector. Wage gaps summarize the wage distribution: they are the ratio of educated workers' to non-educated workers' average wage.

To capture these mechanisms, I build a static one-to-many matching model with transferable utility. Workers and firms differ with respect to their observed characteristics, which are summarized 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. The 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 firm. At equilibrium, wages clear the market and each agent matches with their best option given wages. The model can generate a rich distribution of wages that depend both on workers' and firm's observable characteristics, as well as on the employed workforce. It also predicts equilibrium matching, which is the joint distribution of firms and workforces. Using both matching and wages, I can separately identify firm production from workers' amenities.

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 production parameters over time. This is because by explicitly modeling firms' and workers' match choices, I can 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 depend on a worker's age, education level, and firm sector. Equipped with model predictions for matching and wages, I 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.

The model developed in this paper is related both to one-to-many assignment problems studied in mechanism design (Bikhchandani and Ostroy (2002), Vohra (2011)), and to one-to-one matching models used in family economics (Choo and Siow (2006)). This paper bridges the gap between these two literatures: it extends one-sided assignments to two-sided matching, and generalizes one-to-one matching to one-to-many. Additionally, I extend the econometric framework of Choo and Siow (2006) and Galichon and Salanie (2020) to one-to-many matching.

I use the novel 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, and those who did not. The decrease in wage premium is particularly stark among young workers. 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 imply 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.

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 share 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 start 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 remains 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

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

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 & 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 declined 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 a large number of high school educated workers enter the labor market. First, a decrease in workers' amenities pressures wages upwards. Second, variation in young graduates' share in production compared to more senior high school graduates increases firm demands for the former compared to the latter. I perform several 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 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 firms' demand. These suggest changes in relative productivity in favor of high school graduates have driven the high school wage premium up, but cannot compensate for the large increase in the 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 relies on Gretsky et al. (1992) to show equilibrium existence. This paper is also close to the hedonic model literature (Ekeland et al. (2004), Heckman et al. (2010)). A discussion of the links between hedonic models, matching with transferable utility, and optimal transport can be found in Chiappori et al. (2010). My work is also related to the seminal paper by Kelso and Crawford (1982), 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, which I also use in this paper. I take one-to-many matching models a step further by taking my own model to the data by introducing random shocks that account for unobservables and estimating it. The mechanism design literature has also explored many-to-one assignment problems in a one-sided framework with work by Bikhchandani and Ostroy (2002) and Vohra (2011).

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. Fox (2009) discusses non parametric identification of production functions 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 and agents match on many separate markets. Because static random utility models (including mine) do not follow agents over time, they do not identify the unobserved heterogeneity distribution in the fashion of Abowd et al. (1999), Bonhomme et al. (2019), Bonhomme (2021) and instead focus on match formation based on observable surplus.

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

2 Model

Recent administrative matched employer-employee datasets hold much more information than workers' characteristics and wage. They also inform on firms' characteristics and on matching, i.e. the joint distribution of workers and firms. Besides matching, the data also provides transfers between agents in the form of wage. Relying on this type of dataset 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 and 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 rewrites as a regularized optimal transport problem (Galichon (2016)), and I am therefore able to derive closed-form solutions for predicted matching and wage.

2.1 Setup

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 also 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 workers 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 factor F, meaning that (n, m) and (Fn, Fm) are observationally equivalent. Hence the actual 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 represents 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 characteristics of firm and workers which 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 to 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 their 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 for unobserved heterogeneity. 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 and/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 C. □

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*

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

The mass of type y firms who demand workforce k is

$$D_k^y = m_y \frac{\exp\left(v_{yk}\right)}{\sum_k \exp\left(v_{yk}\right)} \tag{3}$$

Proof. In Appendix C. □

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 large 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 that 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 C ☐

Theorem 2. *Equilibrium matching $\mu_{yk} = D_k^y = \frac{S_{yk}^x}{k_x}$[2] $\forall x$ and equilibrium S_0^x obtain as solution to the social planner problem:*

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

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

$$\begin{aligned} \mathcal{E}(\mu, n, m) = &- \sum_x n_x \sum_y \sum_k \frac{k_x \mu_{yk}}{n_x} \log \frac{k_x \mu_{yk}}{n_x} - \sum_x n_x \frac{S_0^x}{n_x} \log \frac{S_0^x}{n_x} \\ &- \sum_y m_y \sum_k \frac{\mu_{yk}}{m_y} \log \frac{\mu_{yk}}{m_y} \end{aligned}$$

[2]Equality $\mu_{yk} = \frac{S_{yk}^x}{k_x}$ is only defined when $k_x > 0$. If $k_x = 0$, supply S_{yk}^x is not defined

The solution to (5) exists and is unique.

Proof. In Appendix C. □

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_0^x, and second, the first marginal condition $\sum_y \sum_k k_x \mu_{yk} + S_0^x = n_x$ 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_0^x and wages w.

Proposition 3. *Equilibrium matching solves*

$$\begin{aligned} \log \mu_{yk} &= \frac{\Phi_{yk} - \sum_x k_x U_x - V_y + \xi \sum_x k_x \log \frac{n_x}{k_x} + \xi \log m_y}{\xi(1 + \sum_x k_x)} \\ \log S_0^x &= \frac{-U_x + \log n_x}{\xi} \end{aligned} \tag{6}$$

Equilibrium wages write

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

Where U_x, V_y solve

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

Proof. In Appendix C □

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 is increasing in the number of workers hired by the firm. This is reminiscent of search models such as Burdett and Mortensen (1998), although the model presented here is not a search model.

Finally, my 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 works very few hours for instance), we obtain the same intuition as with the representative firm that the marginal change in wage is equal to the 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[3]. 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.

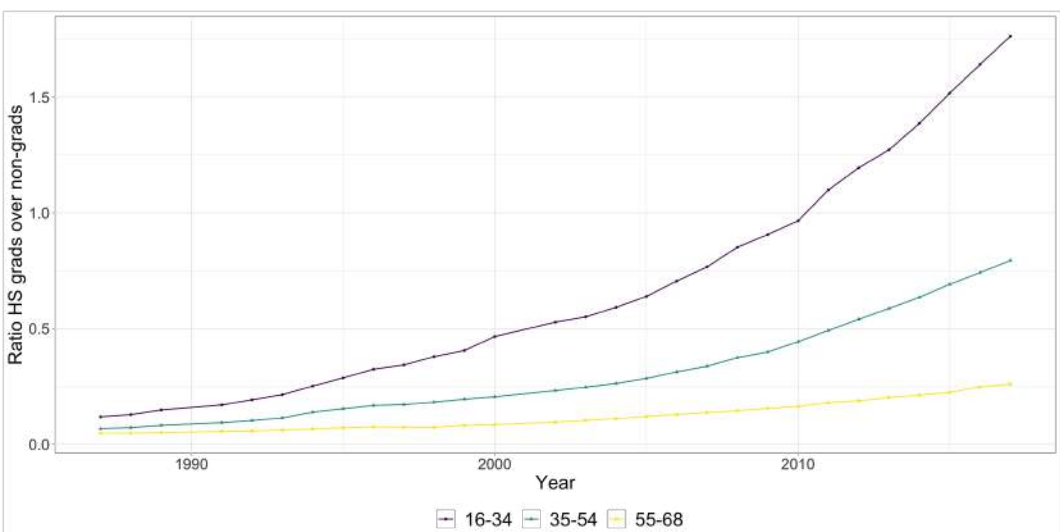
[3]Found on their website

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. Supply of high school graduates relative to non-graduates rises dramatically over the period, as evidenced by Figure 1. Relative supply is measured as the ratio of number of high school graduates over number of active school graduates by age group in each year. Because high school enrolment grows every year, young workers are more 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.

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



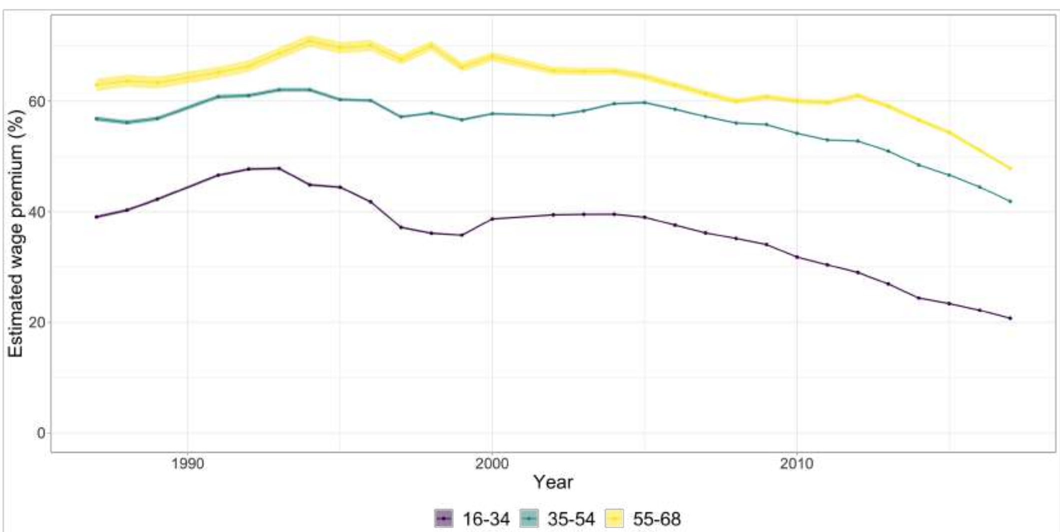
Fact 2 : Wage premia by age group. The second fact that characterizes the Portuguese labor market is the decrease in high school wage premium. To compute high school wage premium by age group, I estimate the following equation by OLS:

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

where each individual i working in firm j at time t earns wage w_{ijt}. a_i is individual i's age group: either y, m or s. $\mathbb{1}_{[\text{HS graduate}_i]}$ equals 1 if i graduated from high school, and 0 otherwise. 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. I allow fixed effects to 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 95% confidence intervals. The high school wage premium differs between age groups: the wage gap is much higher (between 60% and 80% over the period) for senior workers than for younger workers (between 40% and 20%). Figure 2 also shows that the wage premium decreases for all age groups between 1987 and 2017. The extent of the decrease is different depending on age however: 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 by age group



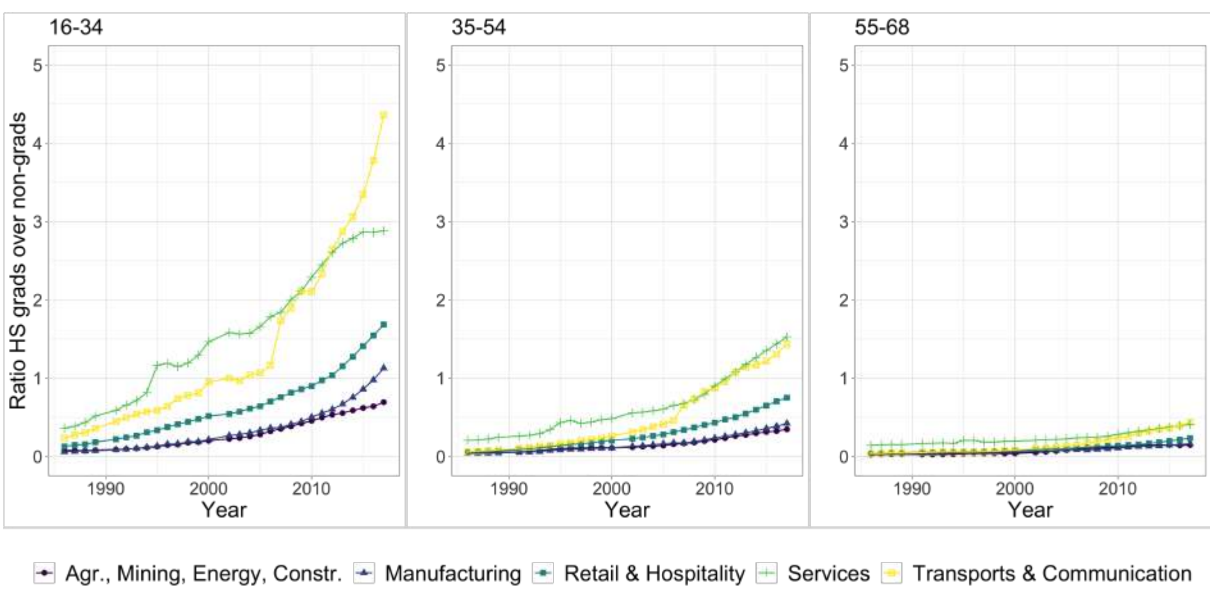
Wage levels differ by gender, with men earning more on average than women in all education levels and age groups (see Figures 9 and 10 Appendix B). Yet, both men and women experience the trend described in Figure 2: the high school wage premium decreases for both genders, more strikingly for young men and women.

Unlike most of the literature, Fact 2 focuses on the high school wage premium, rather than the college wage premium. This choice stems from the particular set up of Portugal:

on the Portuguese labor market, a high school degree is a defining factor in a workers career, because it is less common than in other developed economies. For instance in 2017, 32% of young workers still do not hold a high school degree. However, it is important to note that the university wage premium, defined as the wage gap between university graduates and non graduates, follows a similar trend to the high school wage premium, as shown in Appendix B, Figure 11.

Fact 3 : Sorting between education levels and sector. Sorting between education level and industry is measured by age group as the ratio of the number of employed high school graduates to employed non-graduates in a sector. Sorting is 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 over 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 parametrize the model presented in section 2 to 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

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

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:

$$\alpha_{xyk} = \beta_x^y \tag{11}$$

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:

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

Where υ_{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

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

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[4]. 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, $\mu(\theta, \lambda, \tau, \sigma, \beta)$,

[4]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.

$S_0^x(\theta, \lambda, \tau, \sigma, \beta)$ 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 D.

5 Results

5.1 Parameters estimates

Estimates for high school graduates and non-graduates productivities θ_H^y and θ_L^y by industry y are presented in figure 4. Estimates shows education productivity are heterogeneous by industries, and have evolved in non-linearly: high school graduates prodv knows an impressive surge starting in 2010, especially in the Transport & Communications, Manufacturing and Services industries. Non-graduates productivity drops for all industries between 2010 and 2013. As a result, high school graduates' productivity relative to non-graduates surges at the end of the period.

Figure 4: Estimated education productivities

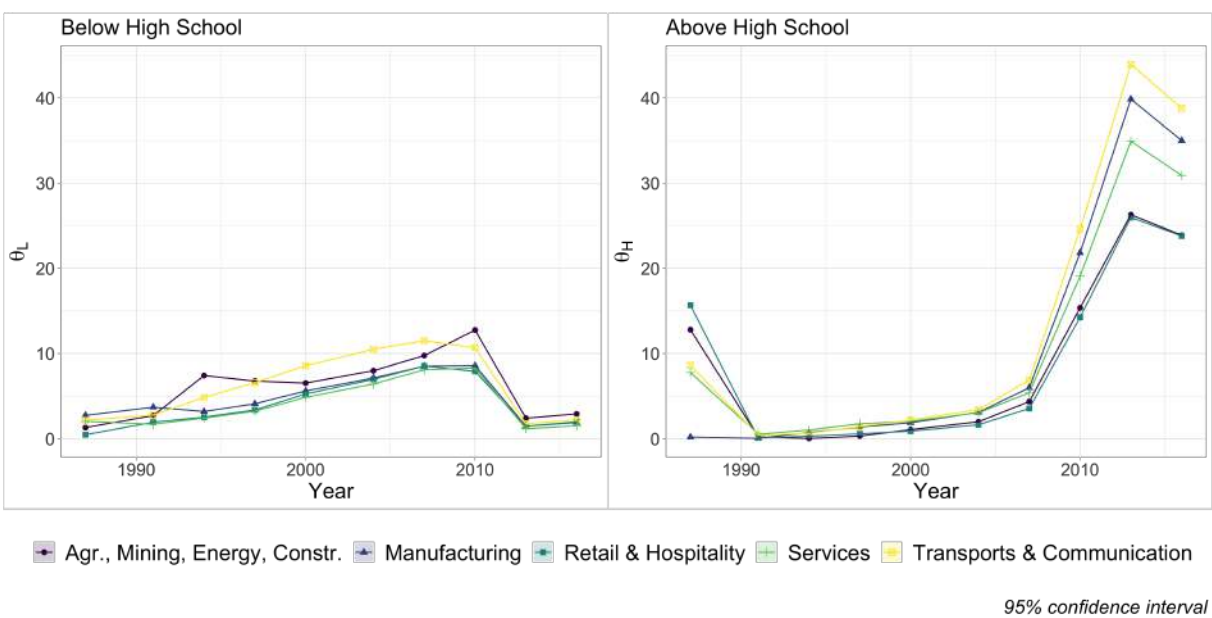


Figure 5 shows the evolution of age productivities $\lambda_{H,a}$ and $\lambda_{L,a}$ by age group and over time. Because $\lambda_{H,a}$ and $\lambda_{L,a}$ sum to one over age groups in any given year, they can be interpreted as shares of each age group in total labor input by high school graduates and non graduates. Estimates $\lambda_{L,a}$ are fairly stable over the period up until 2010-2013, with middle-aged workers making up most of the labor input for non high school graduates. Their share in labor input increases to an even higher level (about 75%) in 2010-2013. 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 at the end of the period.

Figure 5: Estimated age productivities

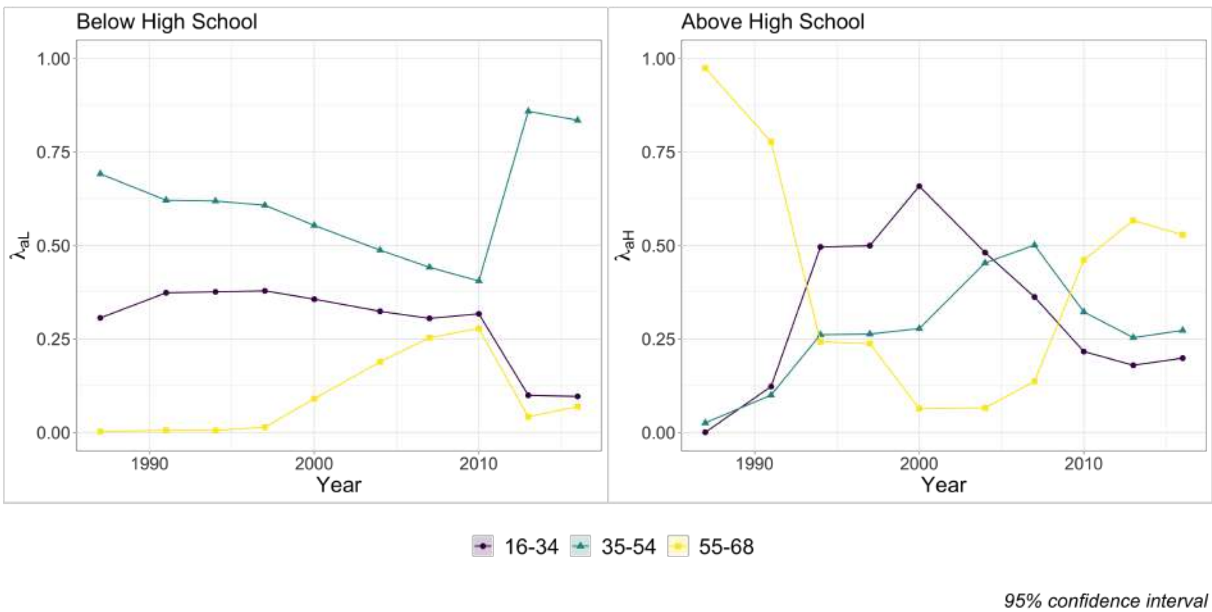
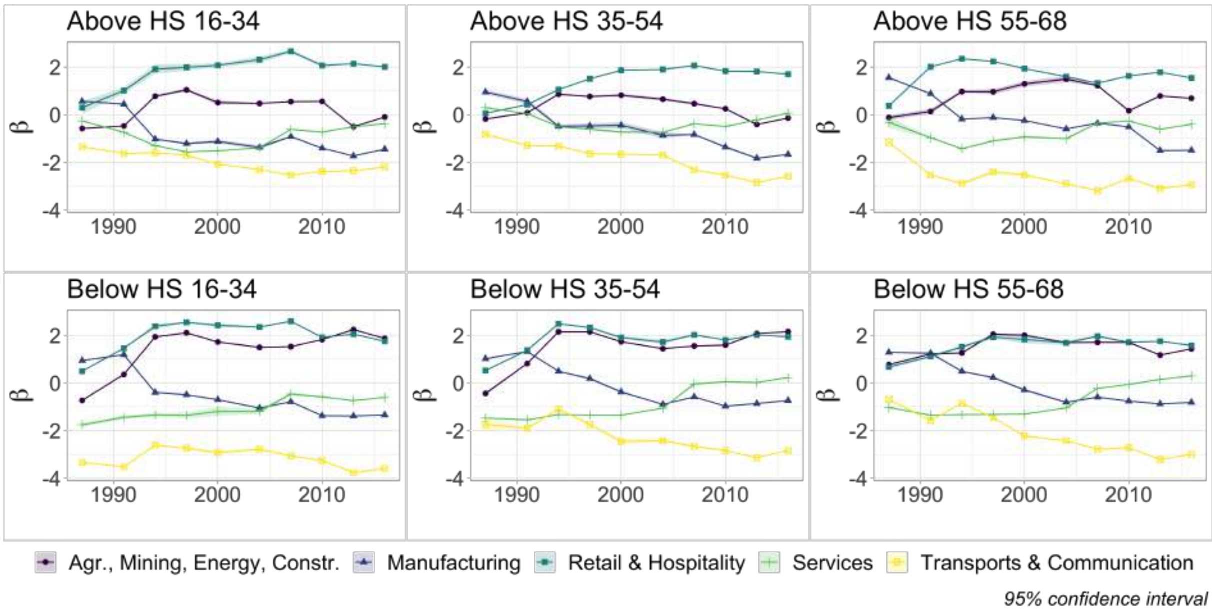


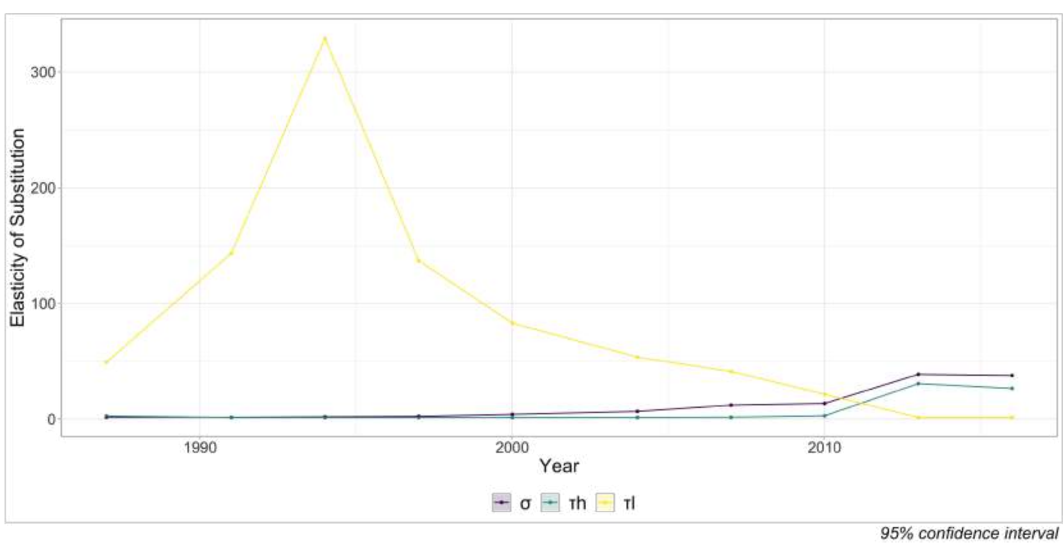
Figure 6 presents the change in worker preferences for firms β_x^y in euros per hours worked. All education levels and age groups hold high preferences for Retail & Hospitality over the period, and low preference for Transport & Communications. High school graduates' preference for Services increases over the period, while their preference for Manufacturing decreases.

Figure 6: Estimated worker preferences



Finally, figure 7 presents estimated elasticities of substitution between education level σ, and age groups τ_H and τ_L. τ_L, the elasticity of substitution between non graduates age groups is generally very high, suggesting age groups are perfect substitutes. σ is between 1.68 in 1987-1989 and increases monotonically to 37.67 in 2016-2017. τ_H is between 2.68 in 1987-1989 and 26.50 in 2016-2017.

Figure 7: Estimated elasticities of substitution



These findings make sense with regards to my estimation method, which unlike most of the literature does not postulate a representative firm by industry, but instead estimates elasticities of substitution at the firm level. Because I account for all firms in the economy and fit the matching distribution, if a large share of firms employs workers of a single education level, either below or above high school, it would be unsurprising to obtain a estimate for σ that suggests that education levels are perfect substitutes (i.e. a very large σ). The same reasoning is valid for substitution between age groups. Scanning the sample for such patterns reveals that a majority of firms employ workers of a single education level: they amount to 78.5% of firms in 1987 and 63.1% in 2017, as well as a single age group: firms who employ a single high school graduates age group amount to 72.8% of firms who employ high school graduates at all in 1987 and 58.2% in 2017. For non graduates, the proportion is between 44.0% in 1987 and 52.0% in 2017. Perfect substitutability of worker types is consistent with the view that the production function at the individual firm level is linear in labor inputs, which is an assumption that had been made in the literature (Hellerstein et al. (1999)).

Discussion. Takeaways from the structural estimates presented in this section are threefold. First, high school graduates productivity has surged over the period. This observation is strongly consistent with a hypothesis of skill-biased technological change, i.e. an increase in worker productivity that favors educated workers, rather than with the competing Hecksher-Ohlin trade hypothesis, which predicts an strengthening of firms' demand for uneducated workers. Second, young and middle-aged high school graduates's share in productivity has increased over the period, which suggests the decreasing high school wage premium for these age groups cannot be explained through an increased demand for experience. Third, workers hold heterogeneous preferences towards sectors. Amenities perceived in the Transport & Communications and Services sector are below zero for most of the period, which puts an upward pressure on wage in these sectors.

These observations must be interpreted in the light of the institutional changes that have occurred in Portugal over the period: the Portuguese labor market is characterized by a steadily (if slowly) increasing minimum wage: in nominal terms, hourly minimum wage is 2.05 euros in 1999 and reaches 3.73 euros in 2017. Most Portuguese workers are also covered by collective bargaining agreements. Finally, it is costly for a firm to fire a worker, because of generous severance packages. Between 2011 and 2014, Portugal has implemented a number of reforms on its labor market: minimum wage was frozen (until the end of 2014), the scope of collective bargaining restricted and terminating workers made less costly. These reforms coincide with a break of trend in the estimated education productivities, and may impact them in two ways. First, the freeze in minimum wage may be partly responsible for the fact that high school non-graduates productivities stop increasing after 2010: because the minimum wage is binding for a large number of non-graduates, its freeze must reverberate on non-graduates' productivity, since the model predicts it increases wage. However, the minimum wage freeze on its own cannot account for the estimated drop in non-graduates productivity, nor the surge in graduates productivity. Both of these appear to be driven by matching, as the number of firms which employ only high school graduates increases rapidly between 2007 and 2013. In light of the institutional changes that have taken place over these five years, changes in matching have been made easier by lowering workers' severance package and reducing the scope of collective bargaining. Hence an interpretation for the surge in graduates productivity is an overdue increase in firm demand that was kept low before 2010-2013 not because of a low productivity, but because of stringent labor market institutions.

5.2 Model Predictions

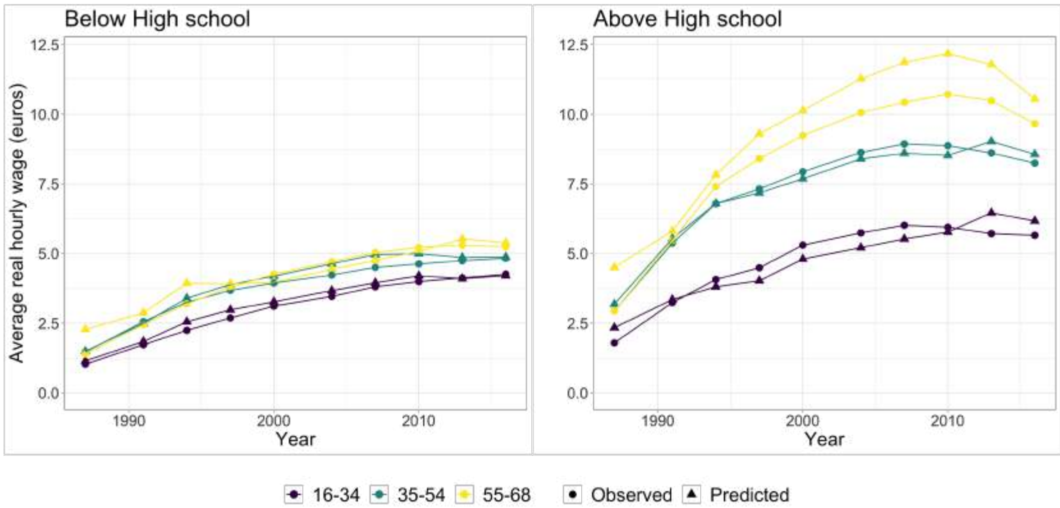
Table 1 compares the slopes of observed and predicted sorting over time. Slopes are obtained by fitting a time trend to the log of relative education supply in each age group and industry. They can be interpreted as average increase in sorting strength (measured as change in relative supply within an 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 quite well, especially for the 16-34 and 35-54 age groups.

Table 1: Sorting average yearly percentage growth - Observed versus Predicted

	16-34		35-54		55-68	
Industry	Data	Prediction	Data	Prediction	Data	Prediction
Agr., Mining, Energy, Constr.	1.274	1.037	0.976	0.876	1.178	0.838
Manufacturing	1.478	1.881	1.222	1.774	0.928	1.341
Retail, Hospitality	1.16	1.274	1.315	1.262	0.933	0.909
Transports, Communication	1.391	1.993	1.748	2.248	1.454	1.947
Services	1.03	1.565	1.05	1.273	0.625	0.728
Overall	0.279	0.25	0.226	0.233	0.181	0.253

Figure 8 compares observed and predicted average wage by education level and age group over time. Model predictions match the slope of average wages for almost all education levels and age groups, except for senior high school graduates. Average wage for this worker type is over-estimated by the model. This is likely due to the importance of collective bargaining in Portugal, which presumably tightens the wage distribution and is not accounted for by the model.

Figure 8: Average wage by education level and age group - Observed versus Predicted



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 ratio of high school graduates over non-graduates employed in a sector y, for a given age group a in a given period t:

$$r(\Gamma^t, \beta^t, n^t, m^t) = \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^t, \beta^t, n^t, m^t)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^t, \beta^t, n^t, m^t)}$$

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

$$\Delta r_{y,a}^{s,t} = \frac{r_{y,a}(\Gamma^s, \beta^s, n^s, m^s)}{r_{y,a}(\Gamma^t, \beta^t, n^t, m^t)}$$

Let t=1987-1989 and s=2016-2017. Then the counterfactual change from labor supply is

$$\Delta r_{y,a}^{LB} = \frac{r_{y,a}(\Gamma^t, \beta^t, n^s, m^t)}{r_{y,a}(\Gamma^t, \beta^t, n^t, m^t)}$$

where $r_{y,a}(\Gamma^t, \beta^t, n^s, m^t)$ is the counterfactual sorting if only labor supply n evolves to its 2016-2017 level, while all other factors Γ, β and m stay at their 1987-1989 levels.

Similarly, define wage premium for age group a in a given period t:

$$\omega(\Gamma^t, \beta^t, n^t, m^t) = \frac{\sum_k k_{H,a}\mu_{yk}(\Gamma^t, \beta^t, n^t, m^t) w_{\{H,a\}yk}(\Gamma^t, \beta^t, n^t, m^t)}{\sum_k k_{L,a}\mu_{yk}(\Gamma^t, \beta^t, n^t, m^t) w_{\{L,a\}yk}(\Gamma^t, \beta^t, n^t, m^t)} - 1$$

so that counterfactual change in wage premium from labor supply is

$$\Delta \omega_{y,a}^{LB} = \frac{\omega_{y,a}(\Gamma^t, \beta^t, n^s, m^t)}{\omega_{y,a}(\Gamma^t, \beta^t, n^t, m^t)}$$

The next tables 2 and 9 show the predicted changes in sorting and wage premium along with the four counterfactual scenarios. Note that because of non-linearities, the sum of changes in all four counterfactuals does not sum to the predicted change.

Table 2 shows changes in relative employment and the increasing presence of high school graduates in the Manufacturing, Services and Transport & Communications sectors are mainly driven by labor supply: the rise in educated workers' share mechanically increases their employment share in each industry. The counterfactual increase in sorting is uniform across industries however, while predicted sorting is not. An important driver of the heterogeneous increase in industries appear to be the evolution of production parameters and worker preferences: production parameters have a particularly strong positive impact on sorting in Manufacturing and Transport & Communications sectors, while worker preferences drive sorting in Transport & Communications and Services to lower levels than other sectors.

Table 2: *Changes in Sorting - Predicted versus Counterfactuals*

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	3.73	9.4	0.87	0.87	0.77
Manufacturing	22.88	8.01	1.01	1.19	0.76
Retail, Hospitality	8.94	8.51	0.92	0.74	0.85
Transports, Communication	43.33	7.04	0.9	4.12	0.55
Services	14.59	6.4	0.79	0.73	0.61

Interpretation: Predicted relative employment of high school graduates to non-graduates is multiplied by 3.73 between 1987 and 2017 in Agr., Mining, Energy, Constr.

Table 9 shows how different scenarios' impact on changes in wage premium. The predicted change in wage premium is a decline in all age groups, especially among young workers. Yet each factor expect the change in industry composition has a heterogeneous effect on wage premium depending on age group. Labor demographics drive young and middle-aged workers wage premium down, as the supply the increase in high school graduates between 16 and 54 makes them less expensive to hire. Surprisingly, the same is not true of senior high school graduates: their supply increase, but so does the supply of senior non-graduates, especially relative to younger non-graduates, so that the change in senior worker wage premium is actually positively impacted by labor demographics. The evolution of production parameters has a positive effect on young workers' wage premium, but a negative effect on other age groups, likely because age productivity of middle-aged non-graduates increases and age productivity of senior graduates decreases over the period. Finally workers preferences have a strong, positive impact on all age groups wage premia. Appendix E shows the details of this impact by industry. For young and middle-aged workers, it appears to be mainly driven by the high school wage premium in the Retail & Hospitality and Services sectors.

Table 3: *Changes in Wage Premium - Predicted versus Counterfactuals*

Age group	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
16-34	-0.28	-1.09	-0.7	0.19	3.67
35-54	-0.18	-0.85	-0.72	-0.28	2.23
55-68	-0.01	6.66	-2.14	-1.31	1.07

Interpretation: the predicted wage premium for the 16-34 age group has fallen by 28% between 1987 and 2017

6 Conclusion

This paper studies wage inequality in Portugal between 1987 and 2017, and seeks to explain the decreasing high school wage premium over the period. The decrease in high school wage premium is particularly stark among young workers, and it is accompanied by a rapidly rising employment share of young educated workers in specific sectors such as Transport & Communications and Services. Over the period, Portugal has experienced a surge in its supply of high school educated workers, as well as sweeping changes in its industry composition, as Services have replaced Manufacturing as the first employer in the country. The increase in educated workers' employment share in the aforementioned sectors suggests a productivity boost in these sectors consistent with skill-biased technological change, but the decreasing high school wage premium observed could also be consistent with a Hecksher-Ohlin theory, whereby less educated workers see their wage rise as they start being more demanded when Portugal joins the European Union;

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 show high school graduates productivity has increased in all sectors, consistent with a theory of skill-biased technological change. Counterfactuals 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.

Chiappori, P.-A., R. J. McCann, and L. P. Nesheim
2010. Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness. *Economic Theory*, 42(2):317–354. Publisher: Springer.

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.

Ekeland, I., J. Heckman, and L. Nesheim
2004. Identification and Estimation of Hedonic Models. *Journal of Political Economy*, 112(S1):S60–S109. Publisher: The University of Chicago Press.

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.

Heckman, J. J., R. L. Matzkin, and L. Nesheim
2010. Nonparametric Identification and Estimation of Nonadditive Hedonic Models. *Econometrica*, 78(5):1569–1591. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA6388.

Hellerstein, J. K., D. Neumark, and K. R. Troske
1999. Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations. *Journal of Labor Economics*, 17(3):409–446. Publisher: University of Chicago Press.

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.

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

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 Details on Empirical Facts

B.1 Wage levels by gender

Figure 9: Average wage by education level and age group - Men

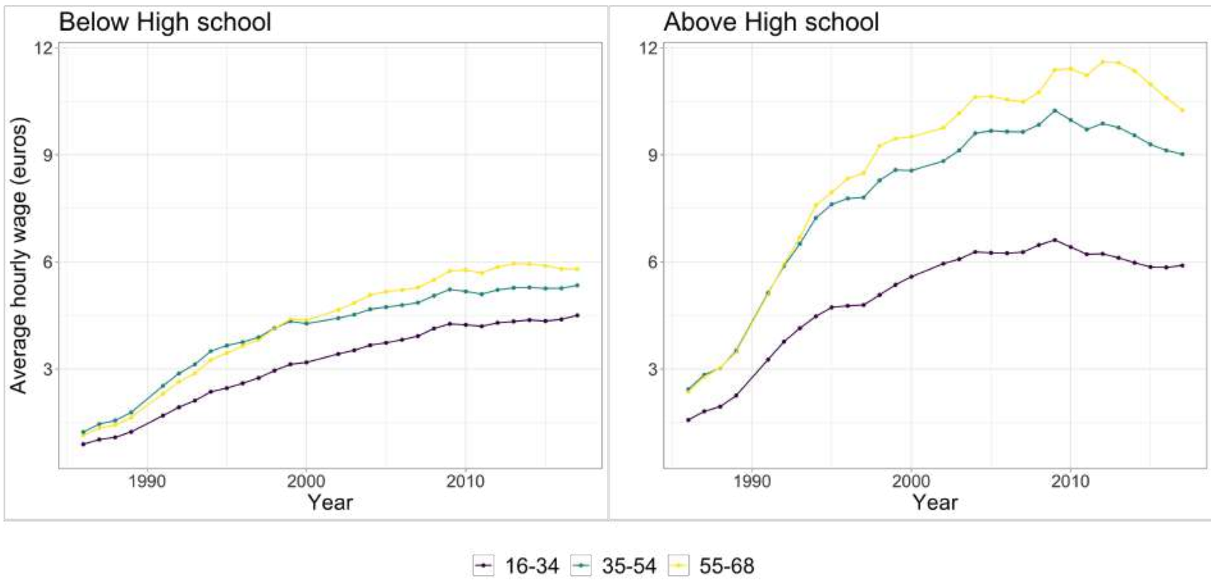
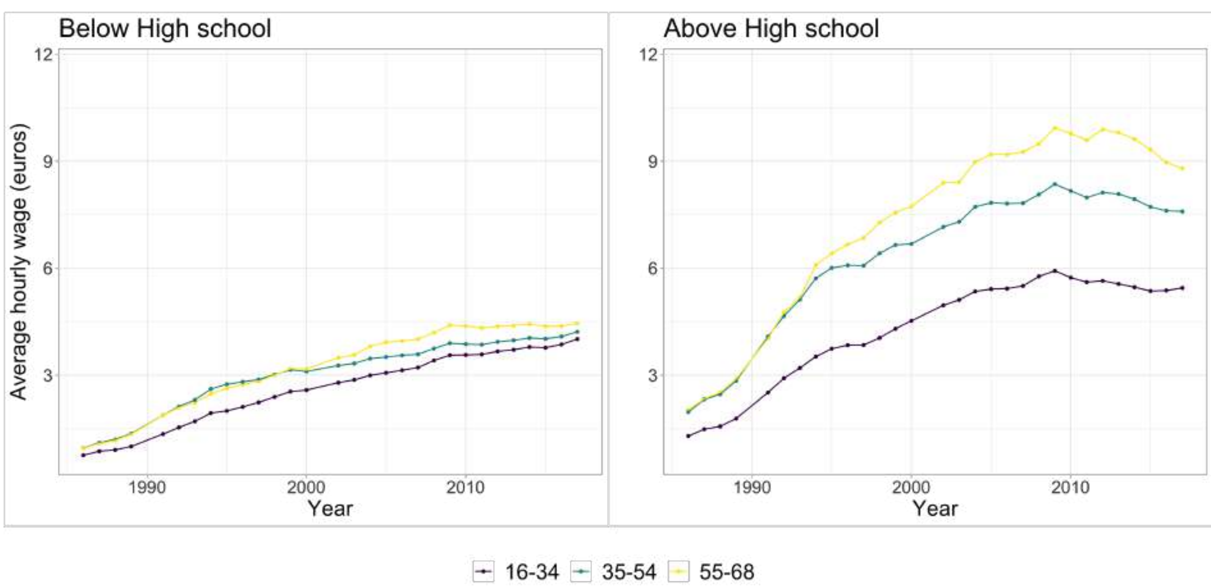


Figure 10: Average wage by education level and age group - Women

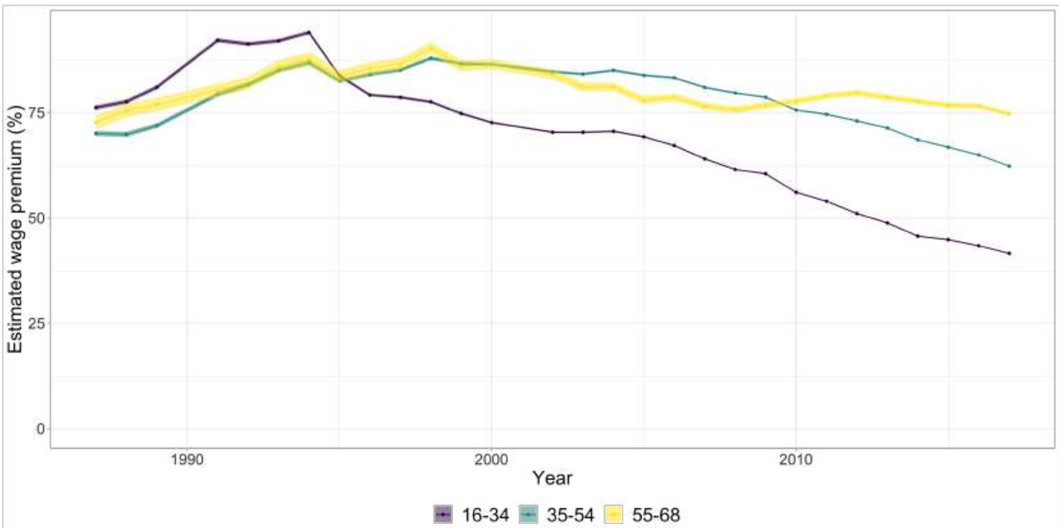


B.2 The university wage premium

The university wage premium is obtained through regressing the following equation:

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

Figure 11: University wage premium by age group



C 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 $\xi \log \sum_k \exp\left(\frac{v_{yk}}{\xi}\right)$.

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

And a similar reasoning shows:

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

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

$\int_j v_j \, dj$:

$$\begin{aligned} &\inf_{u,v} \int_i u_i \, di + \int_j v_j \, dj \\ &\text{s.t } \sum_x \sum_{i:x_i=x}^{k_x} u_i + v_j \geq \Phi_{y_j k} + \xi \sum_x \sum_{i:x_i=x}^{k_x} \epsilon_{i y_j k} + \xi \eta_{jk} \quad \forall k, j \\ &\quad u_i \geq \xi \epsilon_{i0} \end{aligned} \tag{15}$$

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

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

Then (u, v) satisfies (15)'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 (15) becomes $\sum_x k_x u_{xyk} + v_{yk} \geq \Phi_{yk}$.

Applying the law of large numbers, we get that (15) is equivalent to

$$\begin{aligned} &\min_{u,v} \sum_x n_x G_x(u_x) + \sum_y m_y H_y(v_y) \\ &\text{s.t } \sum_x k_x u_{xyk} + v_{yk} = \Phi_{yk} \quad \forall k, y \\ &\quad u_{x0} = 0 \end{aligned} \tag{16}$$

By complementary slackness condition, solving problem (15) 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:

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

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

$$
\begin{aligned}
G_x^*(\mu) &= \max_u \left\{ \sum_{y,k} \frac{k_x \mu_{yk}}{n_x} u_{xyk} + \frac{S_0^x}{n_x} u_{x0} - G_x(u) \right\} \\
H_y^*(\mu) &= \max_v \left\{ \sum_{y,k} \frac{\mu_{yk}}{m_y} v_{yk} - H_y(v) \right\}.
\end{aligned}
$$

For which first order conditions are

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

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

Problem (5)'s objective function is strictly concave and the maximization set defined by the marginal conditions (8) is compact. Therefore the maximum exists and is unique.

$\square$

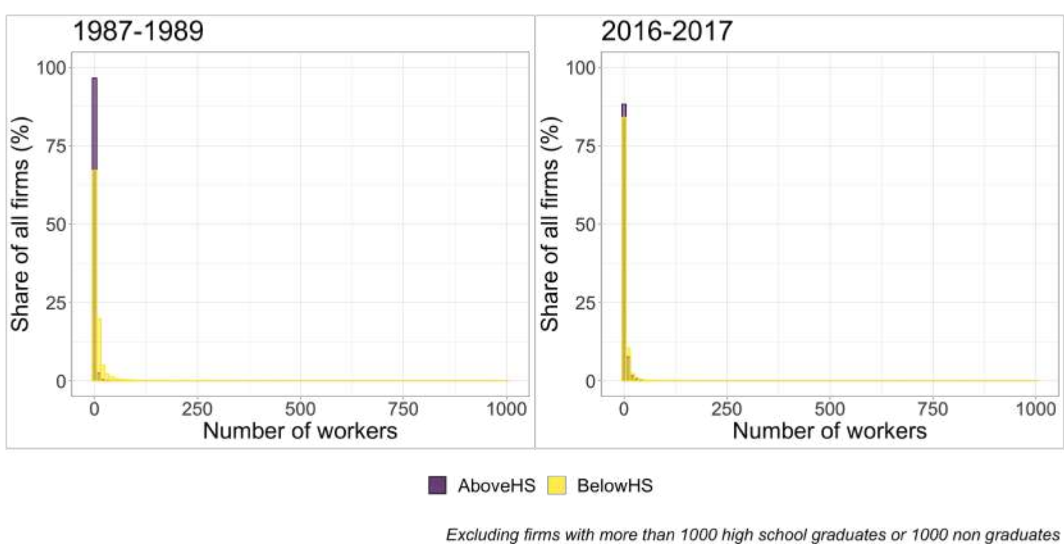
D Estimation

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

D.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 12 in the periods 1987-1989 and 2016-2017. A firm is defined through a distinct firm identifier-year combination. 12 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.

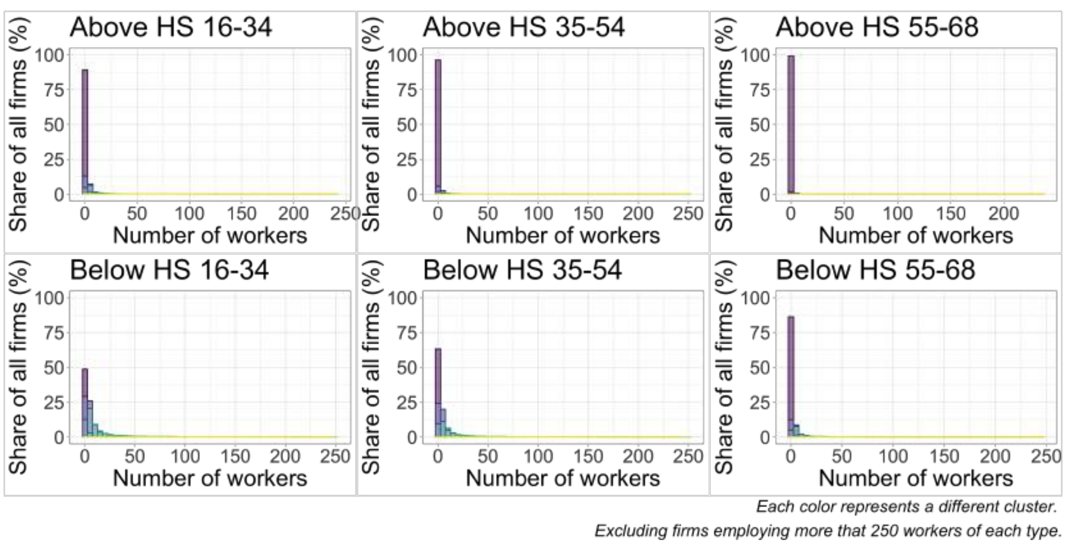
Figure 12: Firm distribution by number of high school graduates and non graduates employed



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 13 displays worker distribution across firms by type, and the clustering of workforce.

Figure 13: Worker type distribution and clusters, 1987-1989



D.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

D.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

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

Where

$$\frac{\partial \log \mu_{yk}(\Gamma, \beta, n, m, s^2)}{\partial \omega} = \frac{1}{1 + \sum_x k_x} \left(\frac{\partial \Phi_{yk}}{\partial \omega} - \sum_x k_x \frac{\partial U_x}{\partial \omega} - \frac{\partial V_y}{\partial \omega} \right)$$

$$\frac{\partial \log S_0^x(\Gamma, \beta, n, m, s^2)}{\partial \omega} = -\frac{\partial U_x}{\partial \omega}$$

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

$\frac{\partial \Phi_{yk}}{\partial \omega}$ 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:

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

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

Which are obtained by differentiating marginal conditions (8).

E Counterfactuals

Table 4: Changes in Sorting in 16-34 age group - Predicted versus Counterfactuals

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	5.18	13	0.81	0.93	0.71
Manufacturing	33.39	12.5	1.03	1.19	0.8
Retail, Hospitality	12.31	10.64	0.9	0.78	0.86
Transports, Communication	44.96	8.09	0.84	1.93	0.46
Services	22.05	9.93	0.74	0.59	0.64

Table 5: Changes in Sorting in 35-54 age group - Predicted versus Counterfactuals

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	4.26	11.51	0.91	0.84	0.81
Manufacturing	27.24	8.78	1.01	1.17	0.7
Retail, Hospitality	10.08	10.45	0.93	0.69	0.81
Transports, Communication	65.55	10.82	0.91	5.89	0.52
Services	18.73	7.82	0.81	0.83	0.59

Table 6: Changes in Sorting in 55-68 age group - Predicted versus Counterfactuals

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	2.87	7.13	1.33	0.59	0.87
Manufacturing	14.69	2.99	0.86	0.93	0.79
Retail, Hospitality	4.43	6.09	0.76	0.3	0.75
Transports, Communication	32.83	5.53	1.39	28.1	0.53
Services	7.27	2.86	1	0.65	0.58

Table 7: Changes in Wage Premium in 16-34 age group - Predicted versus Counterfactuals

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	-0.27	-1.04	-0.64	0.38	-20.76
Manufacturing	-0.56	-1.15	-0.33	0.12	-0.2
Retail, Hospitality	-0.46	-1.03	-0.67	0.14	-4.54
Transports, Communication	0.21	-1.07	-0.52	1.18	0.14
Services	0.71	-1.13	-0.6	0.11	4.47

Table 8: Changes in Wage Premium in 35-54 age group - Predicted versus Counterfactuals

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	-0.2	-0.81	-0.6	-0.24	46.31
Manufacturing	-0.15	-1.02	-0.65	-0.27	0.23
Retail, Hospitality	-0.46	-0.78	-0.69	-0.45	39.02
Transports, Communication	0.11	-0.85	-0.61	0.61	0.18
Services	0.81	-0.88	-0.7	-0.07	3.94

Table 9: Changes in Wage Premium in 55-68 age group - Predicted versus Counterfactuals

Industry	1987-2017 change	Labor supply	Industry composition	Production parameters	Worker preferences
Agr., Mining, Energy, Constr.	-0.49	5.35	-1.9	-1.13	0.29
Manufacturing	0.6	13.56	-4.65	-1.56	1.71
Retail, Hospitality	-0.17	5.76	-2.35	-1.29	0.99
Transports, Communication	-0.15	5.79	-1.4	-0.89	0.42
Services	0.35	4.82	-1.72	-1.39	1.59