Education Expansion, Sorting, and the Decreasing Wage Premium

Pauline Corblet*

October 26, 2021

JOB MARKET PAPER

Click here for the latest version

Abstract

This paper studies the interplay between education expansion and workers and firms sorting in Portugal between 1987 and 2017. The Portuguese labor market is characterized by three facts: a decreasing high school wage premium, a dramatic increase in supply of high school graduates, and an increasingly unbalanced distribution of high school graduates across industries. To quantify the impact of the latter two on the former, I build a model of one-to-many matching where workers sort with firms based on their own preferences, their relative productivity within the firm, and substitution patterns with other workers. Using tool from the optimal transport literature, I solve the model and structurally estimate it on matched employer-employee data. Estimates suggest changes in sorting are mainly driven by heterogeneous increase in relative productivity of high school graduates relative to non graduates across industries. It acts as a mitigating force on the decreasing high school wage premium, but does not fully compensate for high school graduates' rise in relative supply.

*Sciences Po, pauline.corblet@sciencespo.fr
I thank my advisor Alfred Galichon for his valuable guidance and support. I am indebted to Francesco Franco who generously helped me obtain the data used here. I also thank Jean-Marc Robin, Moshe Buchinsky, Florian Oswald, Kerstin Holzheu, Zsofia Barany as well as the participants to Sciences Po PhD and Equiprice seminars for their comments and feedback. All remaining errors are mine.

1 Introduction

It is a well-established fact that high school and university graduates experience higher wages than their less-educated peers in most labor markets. The education wage premium is especially documented for college graduates in the US, where it has increased quickly in the 1970s and 1980s, resulting in a widening of the wage structure and rising inequality between college graduates and their less-educated peers (Goldin and Katz (2008)). Over the same period, the relative supply of college graduates over high school graduates has increased by 4.5% per year between 1967 and 1982 (Card and DiNardo (2002)). The prevailing reason given to the coexistence of these two facts is the rise in college graduates productivity relative to high school graduates', driven by skill-biased technological change (SBTC). SBTC origins in the development of new technologies, in particular computers. 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), Beaudry and Green (2005))

This paper studies the case of Portugal between 1987 and 2017. I highlight two facts on the Portuguese labor market: first the high school wage premium decreases over the period. The high school wage premium is defined as the wage gap between workers who graduated from high school (including workers who pursued higher education), and those who did not. The decrease in wage premium is particularly stark among young workers, and co-occurs with an increase in the supply of high school graduates compared to non-graduates on the labor market. Second, the distribution of high school 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 fact implies relative supply of high school graduates over non-graduates has grown faster than firms relative demand for high school graduates over non-graduates. The latter suggests that sorting between workers and firms has evolved over the period: either because firms in services and transport and communications demand an increasing share of high school graduates, or because high school graduates' preference for these firms strengthens. In the first scenario the impact on high school wage premium in services and transport and communication should be positive, and in the second scenario it should be negative.

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

Motivated by the two facts I evidence for Portugal, decreasing high school wage premium and increasingly unbalanced sorting, I build a static one-to-many matching model with transferable utility to quantify the impact of sorting between workers and firms on the high school wage premium. In the model, workers and firms differ with respect to their observed characteristics, which constitute their type, as well as a stochastic draw of shocks that account for unobserved heterogeneity. A single firm matches with several workers, who constitute its workforce, and surplus created depends on the firms' observable characteristics as well as the workforce's. Utility is transferable under the form of wages paid by the firm to 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 amenities, wage and a random shock. Amenities embody workers' inner preference for a given type of firms. At equilibrium, wages clear the market and each agent match with their best option given wages. The model is able to generate a rich distribution of wages that depend both on worker's and firm's observable characteristics, as well as on employed workforce. It also predicts equilibrium matching, which is the joint distribution of firms and workforces. Using both matching and wages, I am able to separately identify firm production from workers' amenity.

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

I then fit the model 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, 16 to 34, 35 to 54 and 55 to 68 years old. Firms are differentiated by their sector of activity. Following the literature, I choose a nested Constant Elasticity of Substitution (CES) function for production, with productivity parameters for each education level that vary between sectors. I assume worker preferences for firms to depend only on worker's own age and education level, and firm sector. Equipped with predicted matching and wages, I am able to structurally estimate the model on matched employer-employee data. I estimate worker preferences using wage, and firm production parameters using matching, separately every three years.

I find that relative demand for high school graduates from firms in the services and transport and communications has increased over the period, but not enough to absorb the growth in relative supply. I also find that high school graduates' preference for these industries has declined over time.

Related literature. 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 in absorbed on the US labor market by an increased demand for these workers from firms. Card and Lemieux (2001) carry out a similar analysis that further differentiates workers by age, and show that young college graduates are the first to benefit from the slowdown in education attainment in the 1980s. Goldin and Katz (2008) and Autor et al. (2020), among others, relate changes in the US wage structure to be race between education and technology, by which skill biased technological change favors college graduates. 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.

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 borrows tools from the optimal transport literature to introduce unobserved heterogeneity in the form of random utility, and it relies on Gretsky et al. (1992) to show equilibrium existence. Because static random utility models (including mine) do not follow agents over time, they do not identify unobserved characteristics' contribution to match surplus in the fashion of Abowd et al. (1999), and Bonhomme et al. (2019), and instead focus on match formation based on observables. However Fox et al. show that unobserved heterogeneity distribution can be recovered in matching games in which unmatched agents are observed. My work is also related to the seminal paper by Kelso and Crawford (1982) on one-to-many matching, and more recent work by Che et al. (2019) on one-to-many matching with non-transferable utility and Azevedo and Hatfield (2018) on one-to-many matching with transferable utility. They both show existence of equilibrium for a large class of firm preferences, under a large market assumption, an assumption I also use in this paper. Finally, the model I develop is also related to Postel-Vinay and Lise (2015) and Lindenlaub (2017) as it focuses on sorting between multidimensional types.

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

2 Empirical Evidence

2.1 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 unemployment (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.

2.2 Empirical facts

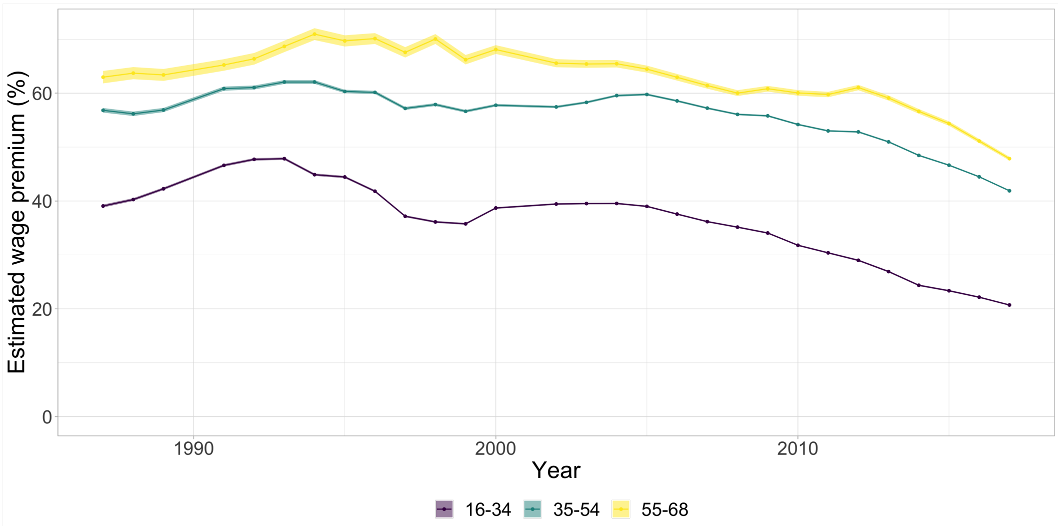
Fact1: Wage premia by age group. To compute high school wage premium by age group, I estimate the following by OLS:

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

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

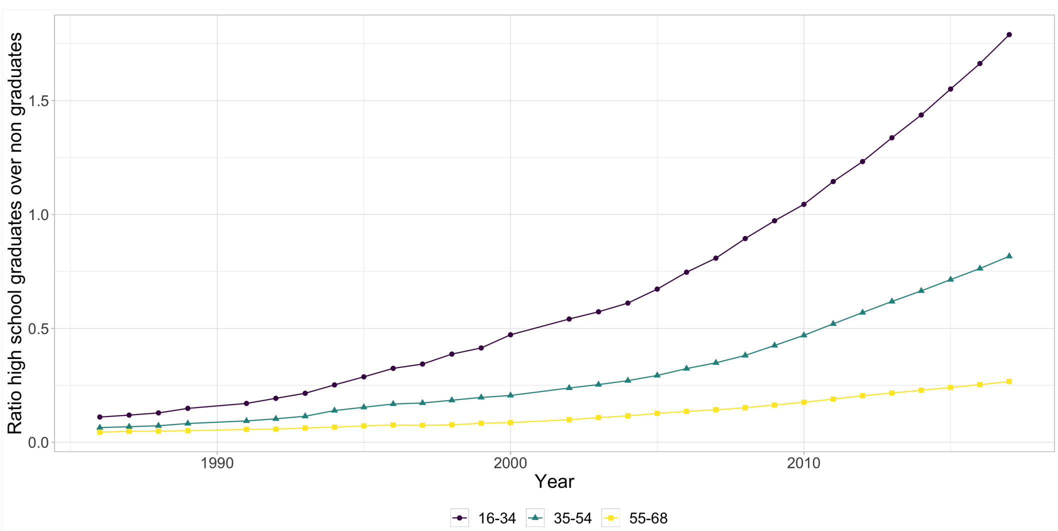
Figure 1 shows the change in estimated high school wage premium over time for each age group, along with 5% confidence intervals. It shows high school wage premia differs by age group: it is much higher (between 60% and 80%) for senior workers than for younger workers (between 40% and 20%). It also shows that the wage premium decreases for all age groups between 1987 and 2017. But the extent of the decrease is different depending on age: senior workers lose only about 17p.p in high school wage premium over the period, while young workers lose slightly less than 50p.p and middle ages workers lose almost 30p.p.

Figure 1: Estimated high school graduate wage premium over non graduates of same age



Relative supply of high school graduates versus non graduates rises dramatically over the period, as evidenced by figure 2. Because high school enrolment grows every year, young workers are most impacted by this growth, and go from .12 to 1.79 on figure 2, meaning high school graduates have grown to be about eight times less numerous to almost twice as numerous as non graduates.

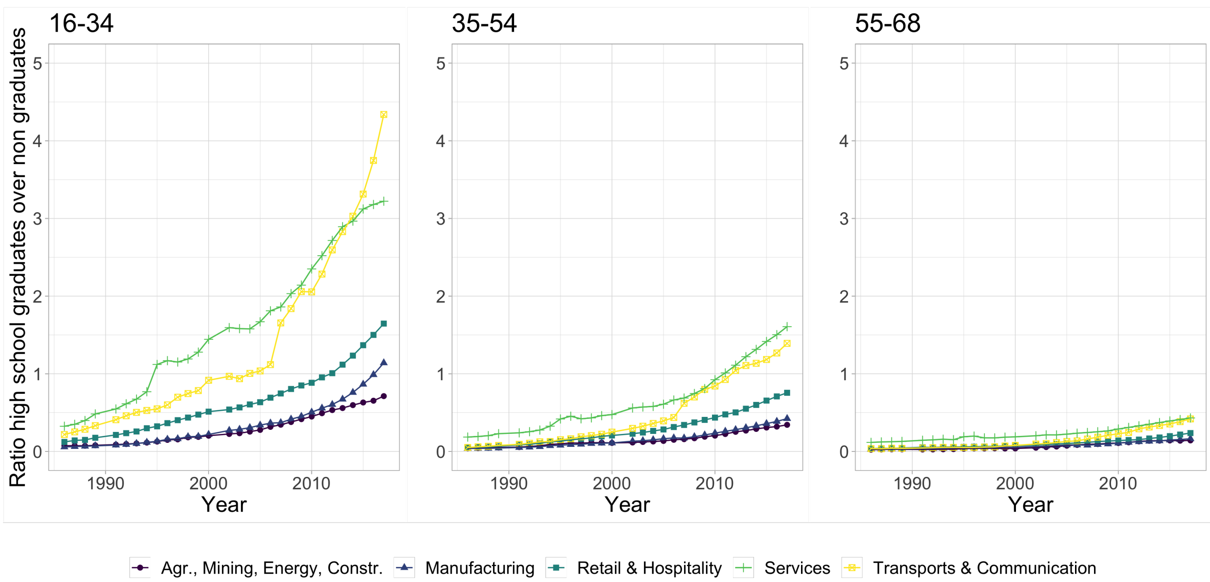
Figure 2: High school graduates versus non graduates relative supply, by age groups



Fact 2: Relative supply and demand by sector. Plotting the ratio of high school graduates to non-graduates by sector reveals stark differences by industry, as shown in figure 3. Most notably, services and transport and communications hire young high school graduates over non graduates at a higher rate than the change relative supply. Relative supply goes

from .11 to 1.79 over the period, while the ratio of graduates to non graduates 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. Comparing high school graduates to non graduates ratios by age groups over the entire market to the same ratio in a given sector yields a measure of sorting. The higher the ratio by sector compared to overall ratio, the stronger the sorting of high school graduates and sector for this age group.

Figure 3: High school graduates versus non graduates relative supply, by age groups and sectors



Summary. 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. Meanwhile, the relative supply of high school graduates over non graduates rises for all age groups, and in particular among young workers. This rise 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. It could be caused either by an increase in relative productivity of high school graduates in these industries, a preference of young high graduates for these sectors, or a rise in substitutability among age groups.

Portugal is unique in that it has known a dramatic education expansion, going from 10% of high school graduates in the labor force in 1987 to about 50% in 2017. It has also known deep changes in how workers sort with firms based on education level, age group, and the firm sector, as evidenced in fact 2. As such, it is an ideal laboratory to understand how sorting between workers and firms drives the high school wage premium overt time. Sorting could change either because of a change in relative supply of workers, a shift in relative productivity in a given sector, changes in workers' preferences over sectors, or variations in the substitution patterns of education levels or age groups. It is likely that all three of these factors impact workers' match with firms between 1987 and 2017 in Portugal, but their impact on high school wage premium differs. In the next section, I build a framework to quantify these changes and evaluate their impact on wage.

3 Model

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

3.1 Set Up

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

Each firm matches with a non negative mass of worker of each type, while each worker matches with a single firm.Let $k_x \in [0, n_x]$ be the mass of type x workers a firm is matched with. Vector k is the workforce employed by the firm, where

$$k = (k_1, \ldots, k_N) \in [0, n_1] \times \ldots \times [0, n_N]$$

Type x worker's utility for being employed at type y firm within workforce k is $u_{xy}(k)$. It is assumed to be additive in a level of amenity α that depends both on worker and firm type, as well as wage w paid by the firm to the worker:

$$u_{xy}(k) = \alpha_{xy}(k) + w_{xyk} \tag{2}$$

Similarly, firm profit is additive in production and total wage paid to its workforce:

$$v_y(k) = \gamma_y(k) - \sum_{x=1}^{N} k_x w_{xyk} \tag{3}$$

I restrict worker amenities functions $(\alpha_{xy})_{x \in \mathcal{X}, y \in \mathcal{Y}}$ and firm productivities functions $(\gamma_y)_{y \in \mathcal{Y}}$ to the set of L^2 functions.

Assumption 1. *$\alpha_{xy}, \gamma_y \in L^2$, for all x, y:*

$$\int |\alpha_{xy}(k)|^2 \, dk < \infty \text{ and } \int |\gamma_y(k)|^2 \, dk < \infty$$

where notation $\int dk$ is short for $\int_0^{n_1} \ldots \int_0^{n_N} dk_N \ldots dk_1$

Total surplus from a match is

$$\Phi_y(k) = \sum_x k_x \alpha_{xy}(k) + \gamma_y(k) \tag{4}$$

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

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

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

$$u_{x_i y}(k) + \epsilon_{iy}(k) \tag{5}$$

Similarly firm j experiences stochastic shock $\eta_j(k)$ in addition to its systematic production:

$$v_{y_j}(k) + \eta_j(k) \tag{6}$$

I impose the following independence conditions on stochastic shocks:

Assumption 2. *Stochastic shocks satisfy the following:*

(i) For each pair of two workers i and i', $\epsilon_{iy}(k)$ and $\epsilon_{i'y}(k)$ are mutually independent and identically distributed

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

(iii) For a worker i and a firm j $\epsilon_{iy}(k)$ and $\eta_j(k)$ are mutually independent

(iV) $\epsilon_{iy}(k)$ is independent of $\alpha_{x_i y}$, $\eta_j(k)$ is independent of γ_y

I further restrict stochastic shocks to the analog of logit shocks in a continuous framework:

Assumption 3.

$$\epsilon_{iy}(k) = \max_n \{\epsilon_i^n : k_i^n = k, y_i^n = y\} \;\; \textit{and}\; \eta_j(k) = \max_n \left\{\eta_j^n : k_j^n = k\right\}$$

Where $(y^n, k^n, \epsilon_i^n)_n$, $(k^n, \eta_j^n)_n$ are enumerations of two inhomogeneous Poisson point processes on of intensity $e^{-\epsilon}$ and $e^{-\eta}$ respectively.

Definition 1. *Expected Utilities*
Define expected utilities on both sides of the market by

$$G_x(u) = \mathbb{E}\left[\max_{y,k}\{u_{xyk} + \epsilon_y(k)\}\right] \;\; \textit{and}\; H_y(v) = \mathbb{E}\left[\max_k\{v_{yk} + \eta(k)\}\right]$$

Under assumption 3, expected utilities rewrite in closed form.

Proposition 1. *If ϵ and η are part of Poisson point processes, then*

$$G_x(u_x) = \log \sum_y \int \exp(u_{xy}(k))\, dk \text{ and } H_y(v_y) = \log \int \exp(v_y(k))\, dk$$

Proof. In appendix B □

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. I describe the competitive equilibrium in the next section.

3.2 Competitive Equilibrium

Loosely speaking, the equilibrium on a market is found when supply from workers meets demand from firms. Supply and demand should also be feasible: workers should not ask for more firms y than m_y, nor should firms ask for more than n_x workers of type x. I formally define supply and demand below.

Definition 2. *Supply*
Define supply $S = (S^x)_{x\in\mathcal{X}}$ where $S_y^x(k)$ is the mass of type x workers willing to match with type y firm and mass k workforce and S_0^x is the mass of type x workers willing to remain unmatched. Supply over all firm types and workforces must add up to n_x:

$$\sum_y \int S_y^x(k)\, dk + S_0^x = n_x\, \forall x$$

S is feasible iff $S \in \mathcal{S} = \left\{ (S^x)_x \;\middle|\; \int \frac{S_y^x(k)}{k_x}\, dk \leq m_y\, \forall x,\, \forall y \right\}$[2]

Definition 3. *Demand*
Define demand $D = (D^y)_{y\in\mathcal{Y}}$ where $D^y(k)$ is the mass of type y firms willing to match with workforce k. Demand for all workforces must add up to the mass of firms of type y:

$$\int D^y(k)dk = m_y\, \forall y$$

D is feasible iff $D \in \mathcal{D} = \left\{ (D^y)_y \,, \;\middle|\; \sum_y \int k_x D^y(k)\, dk \leq n_x\, \forall y \right\}$

In the context of one-to-many matching, supply S and demand D are measured in different 'units': if a firm can match with numerous workers types, workers can only match with one

[2]To avoid integrability issues in feasibility condition, consider the function defined as $\frac{S_y^x(k)}{k_x}$ if $k_x > 0$ and 0 if $k_x = 0$

firm type. This shows in the feasibility conditions for S and D: a mass $S_y^x(k)$ of type x workers supplying firms of type y within workforce k corresponds to $\frac{S_y^x(k)}{k_x}$ firms supplied. However a mass $D^y(k)$ demanding k_x workers corresponds to $k_x D^y(k)$ workers demanded. These equivalences are translated into feasibility conditions for both S and D

I model unemployment through S_0^x, which is determined at equilibrium. I assume no counterpart on the firm side: all firms must be matched to a given workforce.

The difference between supply and demand at the worker type, firm type, and workforce level is defined as excess demand.

Definition 4. *Excess Demand*
Given types x, y and workforce mass k, excess demand is defined as

$$Z_{xyk}(S, D) = k_x D^y(k) - S_y^x(k)$$

A competitive equilibrium is reached on the market when supply and demand are feasible, excess demand is zero, and matching is incentive compatible, in the following sense:

Definition 5. *Competitive Equilibrium*
An arrangement $[S, D, w]$ is a competitive equilibrium if it satisfies:

(i) S and D are feasible: $S \in \mathcal{S}$ and $D \in \mathcal{D}$

(ii) S and D clear the market: $\forall x \in X$, $y \in Y$, $k \in \mathcal{K}$: $Z_{xyk}(S, D) = 0$

(iii) Given w and individual draws of ϵ and η, each agent obtains their optimal utility/profit:

$$\forall i \in I, S_{y^*}^{x_i}(k^*) > 0 \Rightarrow (y^*, k^*) \in \arg\max_{y,k} u_{x_i y}(k) + \epsilon_{iy}(k)$$
$$\forall j \in J, D^{y_j}(k^*) > 0 \Rightarrow k^* \in \arg\max_{k} v_{y_j}(k) + \eta_j(k)$$

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.

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

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

$$\begin{aligned}
&\inf_{u,v} \sum_x n_x G_x(u) + \sum_y m_y H_y(v) \\
&s.t \sum_x k_x u_{xy}(k) + v_y(k) = \Phi_y(k) \quad \forall k, y \\
&\quad u_{x0} = 0
\end{aligned} \tag{7}$$

Proof. In appendix B □

Theorem 2. *Equilibrium supply and demand* $\frac{S^x_y(k)}{k_x} = D^y(k)$ *obtain as solution to the social planner problem:*

$$\begin{aligned}
&\max_{\mu, S_0} \sum_y \int \Phi_y(k)\mu_y(k)\, dk + \mathcal{E}(\mu, n, m) \\
&s.t \sum_y \int k_x \mu_y(k)\, dk + S_0^x = n_x \\
&\quad \int \mu_y(k)\, dk = m_y
\end{aligned} \tag{8}$$

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

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

The solution to (8) *exists and is unique.*

Proof. In appendix B □

Theorem 2 shows that equilibrium matching can be obtained by solving a penalized social planner problem, where the objective function is the difference between total expected surplus and an entropy term due to unobserved heterogeneity. It is reminiscent of the regularized optimal transport problem with one dicrete and one continuous marginal (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 $\sum_y \int k_x \mu_y(k)\, dk + S_0^x = n_x$ is not a condition on the marginal distribution of k, which is endogeneous here, but on the marginal distribution if worker types.

Proposition 2. *Equilibrium matching solves*

$$\begin{aligned}\log \mu_{yk} &= \frac{\Phi_y(k) - \sum_x k_x U_x - V_y + \sum_x k_x \log \frac{n_x}{k_x} + \log m_y}{1 + \sum_x k_x} \\ \log S_0^x &= -U_x + logn_x\end{aligned} \tag{9}$$

Equilibrium wages write

$$\begin{aligned} w_{xy}(k) =& \frac{\gamma_y(k) - \alpha_{xy}(k) + U_x - V_y + \log m_y - \log \frac{n_x}{k_x}}{1 + \sum_x k_x} \\ &+ \frac{\sum_{x' \neq x} k_{x'} \left((\alpha_{x'y}(k) - \alpha_{xy}(k)) - (U_{x'} - U_x) + \log \frac{n_{x'} k_x}{n_x k_{x'}} \right)}{1 + \sum_x k_x} \end{aligned} \tag{10}$$

Where U_x, V_y solve

$$\begin{cases} \sum_y \int k_x \exp\left(\frac{\Phi_{yk} - \sum_x k_x U_x - V_y + \sum_x k_x \log\left(\frac{k_x}{n_x}\right) + \log m_y}{1 + \sum_x k_x} \right) dk = n_x \\ \int \exp\left(\frac{\Phi_{yk} - \sum_x k_x U_x - V_y + \sum_x k_x \log\left(\frac{k_x}{n_x}\right) + \log m_y}{1 + \sum_x k_x} \right) dk = m_y \end{cases} \tag{11}$$

Proof. In appendix B ☐

3.3 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 semi-discrete one-to-one framework, but for the specific shape of marginal conditions in (11), that links the matching over workforces and firms back to the number of workers of each type. Unlike in Choo and Siow (2006) however; I observe tranfers as wages and can leverage them to split total match surplus between workers and firms, in the spirit of Dupuy and Galichon (2017)

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

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

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

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

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

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

4 Production Framework

Each firm on the labor market produces according to a nested Constant Elasticity of Substitution (CES) production function. Each firm belongs to some type y, which is observed and determines the parameters of its production function. Firm production is a function of labor input by education level and age bin in any given year t:

$$\gamma_y^t(H, L) = \left[\left(\theta_H^{y,t} H\right)^{\frac{\sigma^t - 1}{\sigma^t}} + \left(\theta_L^{y,t} L\right)^{\frac{\sigma^t - 1}{\sigma^t}}\right]^{\frac{\sigma^t}{\sigma^t - 1}} \tag{12}$$

Production γ's outer nest involves three parameters that may all vary with time: $\sigma(t)$, $\theta_H^y(t)$, $\theta_L^y(t)$ and two aggregate inputs $H(t)$ for high school graduates and $L(t)$ for non high school graduates. $\sigma(t) \in (0, \infty)$ is the elasticity of substitution between education levels, it is greater than one if graduates and non graduates are gross substitutes, and smaller than one if they are gross complements. $\sigma(t)$ is assumed to be the same across firm types. $\theta_H^y(t)$, $\theta_L^y(t) \in [0, \infty)$ are graduates and non graduate's productivity parameters. Both parameters may vary by firm type y. Aggregate labor inputs $H(t)$ and $L(t)$ are defined as

$$H = \left[\sum_{a \in \{y,m,s\}} \lambda_{a,H}^{y,t} H_a^{\frac{\tau_H^t - 1}{\tau_H^t}} \right]^{\frac{\tau_H^t}{\tau_H^t - 1}}$$
$$L = \left[\sum_{a \in \{y,m,s\}} \lambda_{a,L}^{y,t} L_a^{\frac{\tau^L(t) - 1}{\tau^L(t)}} \right]^{\frac{\tau_L^t}{\tau_L^t - 1}} \tag{13}$$

Where H_a and L_a) are labor inputs provided to the firm by high school graduates and non graduates in age category a. Labor input is measured in number of full-time equivalent workers. $\lambda_{a,H}^{y,t}$, $\lambda_{a,L}^{y,t} \in [0, \infty)$ are age productivity parameters, differentiated by education level and firm type. τ_H, $\tau_L \in (0, \infty)$ are elasticities of substitution between age categories. They vary by education level but are the same across firm types.

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 $\theta_H^{y,t}$ and $\theta_L^{y,t}$. Unbiased technological change that increases all workers productivity results in an increase in both $\theta_H^{y,t}$ and $\theta_L^y y, t$. 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:

$$\sum_a \lambda_{a,H}^{y,t} = \sum_a \lambda_{a,L}^{y,t} = 1 \quad \forall y \tag{14}$$

5 Identification and Estimation

The model's predictions on matching (9) and wage (10) 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. One must assume that amenities are zero in order to identify production, as in the reduced forms in section 7.

I am aiming to parametrically estimate α and γ. The functional forms are the same as the ones laid out in section 4: there are $N = 6$ worker types that are the combination of two education levels, high school graduates and non graduates, and three age groups, young (below 35), middle-aged (between 35 and 54), and senior (above 55). Let $e(x)$, $a(x)$ be type x's education level and age group. Firm workforce $k(t)$ at time t is composed of employed masses of each worker types:

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

Employed mass of worker k_x is 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. I allow all amenity and production parameters to vary over time. and Type y firm production $\gamma_y^t(k)$ at time t is

$$\gamma_y^t(k) = \left[(\theta_H^y(t) H(t))^{\frac{\sigma(t)-1}{\sigma(t)}} + (\theta_L^y(t) L(t))^{\frac{\sigma(t)-1}{\sigma(t)}} \right]^{\frac{\sigma(t)}{\sigma(t)-1}}$$

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

$$\begin{aligned} H(t) &= \left[\sum_{a \in \{y,m,s\}} \lambda_{a,H}(t) k_{H,a}(t)^{\frac{\tau^H(t)-1}{\tau^H(t)}} \right]^{\frac{\tau^H(t)}{\tau^H(t)-1}} \\ L(t) &= \left[\sum_{a \in \{y,m,s\}} \lambda_{a,L}(t) k_{L,a}(t)^{\frac{\tau^L(t)-1}{\tau^L(t)}} \right]^{\frac{\tau^L(t)}{\tau^L(t)-1}} \end{aligned} \tag{15}$$

I assume worker amenities are constant in k:

$$\alpha_{xy}^t(k) = \beta_x^y(t) \tag{16}$$

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 $\left(\lambda_{a,H}^y(t)\right)_a$, $\left(\lambda_{a,L}^y(t)\right)_a$, $(\theta_H^y(t))_y$, $(\theta_L^y(t))_y$, $(\beta_x^y(t))_{x,y}$, $\tau_H(t)$, $\tau_L(t)$ and σ. The first step estimates amenities parameters and the second production parameters.

Step 1: I estimate amenity parameters $(\beta_x^y(t))_{x,y}$ through equation (10). Assume $\tilde{w}_{ij}$, worker i's wage for working at firm j is observed in the data with measurement error, i.e. for any individual worker i of type x employed firm j of type y within workforce k:

$$\tilde{w}_{ij} = w_{xy}(k) + \epsilon_{ij} \text{ where } \epsilon \sim \mathcal{N}(0, \kappa_j s^2) \text{ iid} \tag{17}$$

Where κ_j is a constant term that depends on firm j. Given the previous assumption, average wage paid to type x workers in firm j that employs workforce k is normally distributed with mean 0 and variance $\kappa_j s^2$.

Let x^* be any reference worker type. Then differentiating equation (10) for a given worker type x, workforce k and industry y obtains:

$$w_{x^*y}(k) - w_{xy}(k) = U_{x^*} - U_x + \beta_x^y - \beta_{x^*}^y + \log\left(\frac{n_x k_{x^*}}{n_{x^*} k_x}\right) \tag{18}$$

Where production $\gamma_y(k)$ has been differentiated out.

In order to use equation (19) to estimate β_x^y, I choose reference type x^* to be non high school graduates between 16 and 34. and I make the following assumption:

Assumption 4. *Reference type x^* if indifferent between firm types, and has null amenity parameters: $\beta_{x^*}^y = 0$*

Under the previous assumption, equation (19) at firm j's level becomes

$$\begin{gathered} \bar{w}_{x^*j} - \bar{w}_{xj} - \log\left(\frac{n_x k_{j,x^*}}{n_{x^*} k_{j,x}}\right) = (U_{x^*} - U_x + \beta_x^{y_j}) + \epsilon_{j\kappa} \\ \text{where } \epsilon_{xj} \sim \mathcal{N}(0, \kappa_j s^2) \text{ iid} \end{gathered} \tag{19}$$

Where $\bar{w}_{x^*j}$,$\bar{w}_{xj}$ are average wages paid by firm j to types x^* and x. Hence amenity parameters can be estimated with Generalized Least Squares through

$$y_{xj} = a_x + b_{xy} + \epsilon_{xj} \tag{20}$$

Where $y_{xj} = \bar{w}_{x^*j} - \bar{w}_{xj} - \log\left(\frac{n_x k_{x^*j}}{n_{x^*} k_{xj}}\right)$. b_{xy} is identified through of variation in k. There is however some colinearity in the equation, due to the presence of a_x, which leads to the following additional assumption:

Assumption 5. $\beta_x^{y^*} = 0$ *for some firm type* y^*

I choose y^* to be the Agriculture, Mining, Energy and Construction industry.

Step 2: In a second step, I estimate remaining parameters $\left(\lambda_{a,H}\right)_a$, $\left(\lambda_{a,L}\right)_a$, $\left(\theta_H^y\right)_y$,$\left(\theta_L^y\right)_y$, σ, τ_h, τ_l through maximum likelihood on matching. Let $\tilde{\mu}_{yk}$ be the observed mass of type y firms matched with type k workforce. Let μ_{yk} be the prediction (9) made by the model. In particular μ_{yk} satisfies margin equations (11). The log likelihood is:

$$\sum_{y,k} \tilde{\mu}_{yk} \log \mu_{yk}$$

Maximizing the log likelihood requires to solve

$$\max_{\sigma,\tau_l,\tau_h,\lambda_{a,H},\lambda_{a,L},\theta_H^y,\theta_L^y} \sum_{y,k} \tilde{\mu}_{yk} \left(\Phi_{yk} - \sum_x k_x U_x - V_y + \sum_x k_x \log\left(\frac{k_x}{n_x}\right) + \log m_y\right) \quad \text{s.t } U, V \text{ statisfy (11)} \tag{LL}$$

Given values for $(\beta_x^y)_{x,y}$, productivity parameters $(\theta_h^y)_y$, $(\theta_l^y)_y$, $\left(\lambda_{a,H}(t)\right)_a$, $\left(\lambda_{a,L}(t)\right)_a$, and elasticities of substitution σ, τ_h, τ_l are identified through first order conditions to problem (LL).

I run both estimation steps on ten separate three year periods between 1987 and 2017[3]. I exclude from estimation firms who employ only one worker type, as the CES production function is not well defined for these firms. These represent 6.3% of workers over the entire period.

[3] Since data for years 1990 and 2001 are missing, the last time period spans only two years.

6 Results

6.1 Parameters estimates

Estimates for $\frac{\theta_H^y}{\theta_L^y}$ by industry y are presented in figure 4. They show evidence of heterogeneity in relative demand by industry: for primary industries and manufacturing, $\frac{\theta_H^y}{\theta_L^y}$ is below 1 over the entire period, meaning high school graduates' productivity in these industries lower than non graduates'. On the contrary, services and transport and communications ratio is in favor of high school graduates. Services in particular display a sustained increase in relative productivity until the early 2000s, but the ratio decreases in the late 2000s and stabilizes slightly below 3, meaning high school graduates are still almost three times as productive as non graduates.

Figure 4: Estimated education productivities ratio

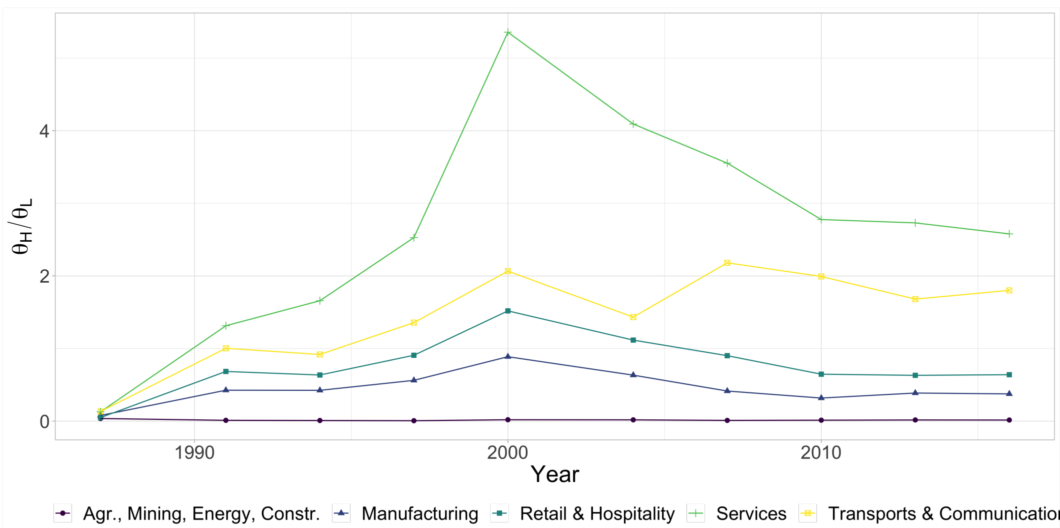


Figure 5 shows the evolution of age productivities $\lambda_{H,a}$ and $\lambda_{L,a}$ by age group and over time. Because $\lambda_{H,a}$ and $\lambda_{L,a}$ sum to one over age groups in any given year, they can be interpreted as shares of each age group in total labor input by high school graduates and non grdauates. Young workers' productivity increases and middle-aged workers' decreases for both education levels. Senior workers' productivity is lower than the other two age groups for both education levels. It does increase to about 20% for high school graduates towards the end of the period.

Figure 5: Estimated age productivities

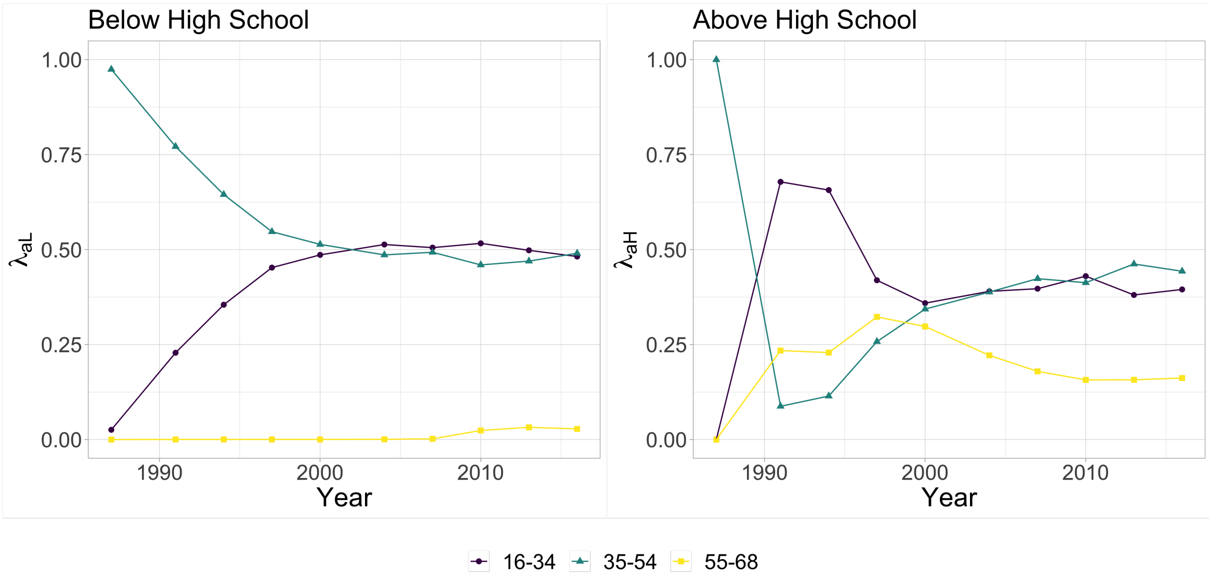
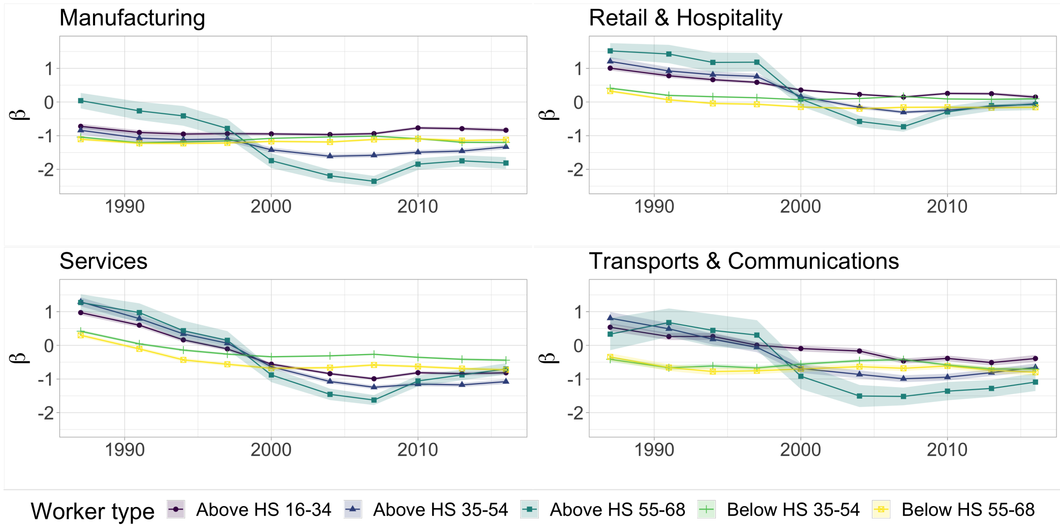


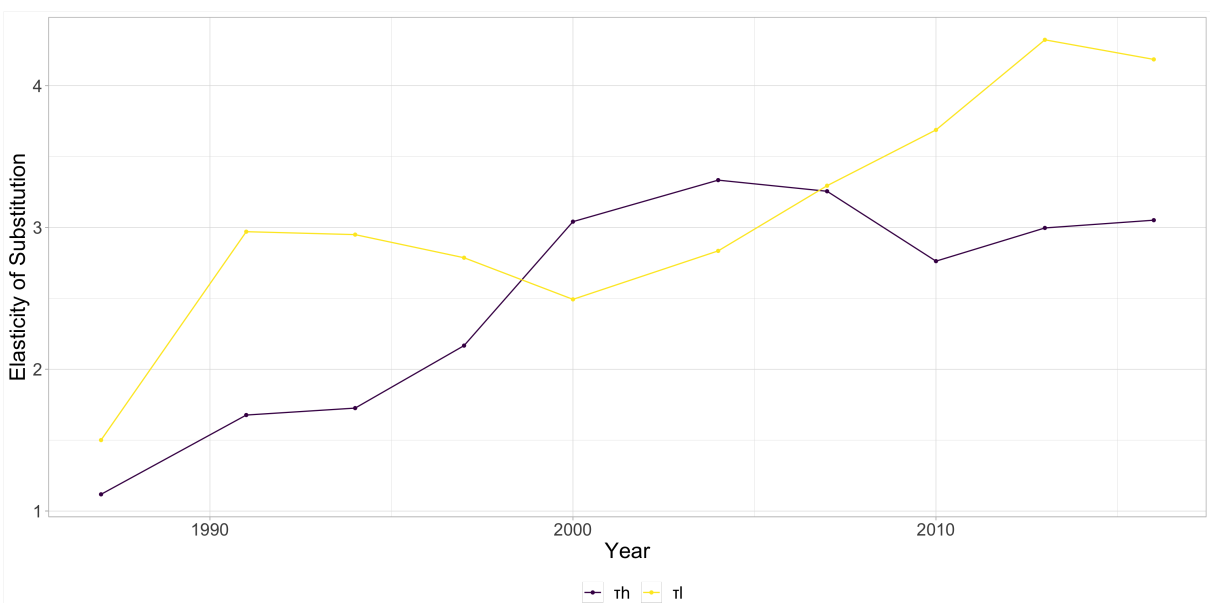
Figure 6 presents the change in worker preferences for firms β_x^y in euros per hours worked. It is computed with respect to reference industry agriculture, mining, energy, construction, and reference workers between 16 and 34 with non high school degree. All education levels and age groups display a disutility for working in manufacturing. High school graduates of all age groups see their preference for services decrease over the period, while non graduates' preferences remain fairly constant. Amenities in retail and hospitality and transport and communications remain constant and positive for all worker types except experienced high school graduates. This group is the one that experiences the largest negative changes in their preferences towards sectors

Figure 6: Estimated worker preferences



Finally, figure 7 presents estimated elasticities of substitution between ages. It shows substitution between age groups is generally higher among non high school graduates than graduates, and both elasticities are increasing over the period.

Figure 7: Estimated elasticities of substitution



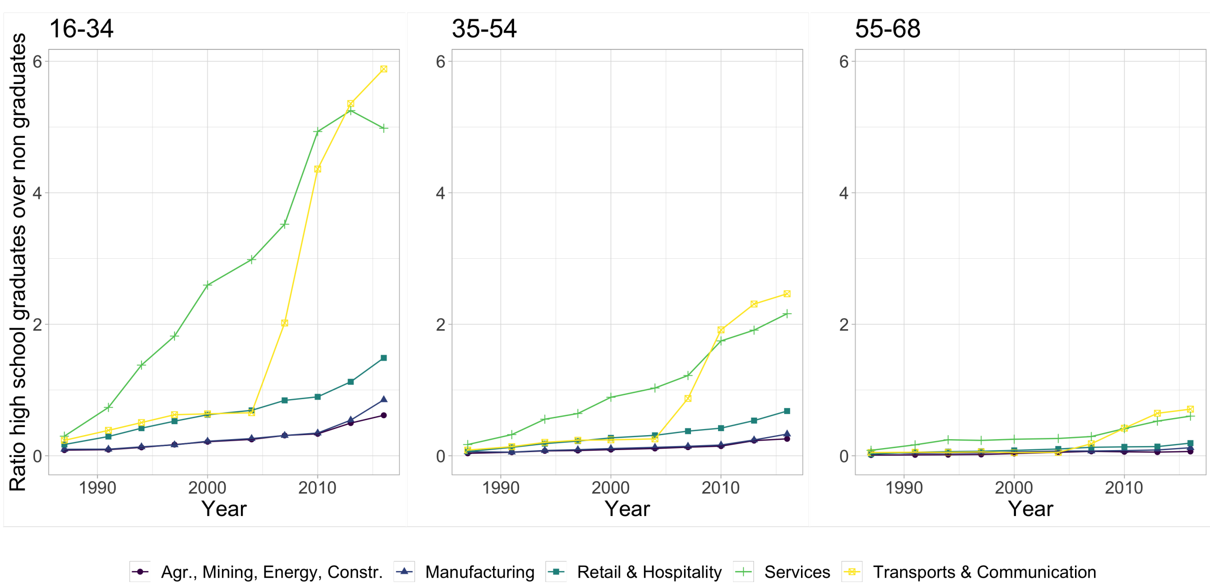
I find estimates for σ, the elasticity of substitution between education levels, between 425 in 1987 and 161 in 2016. In other words, high school graduates and non graduates are almost perfectly substitutable in Portugal over the period. This finding makes sense with regards to my estimation method,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

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

6.2 Predicted sorting

Figure 8 presents the analogous to figure 3, given the parameters predicted by the model. It matches figure 3 well, except for an over estimation of the strength of sorting between young high school graduates and the services and transport and communications sectors.

Figure 8: Predicted sorting



7 Comparison to Card and Lemieux (2001)

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

$$\max_{H_a, L_a} \gamma(t) - \sum_{a \in \{y,m,s\}} H_a w_{H,a} - \sum_{a \in \{y,m,s\}} L_a w_{L,a} \tag{21}$$

Where $\gamma(t)$ is the CES production function described in section 4 with no dependence on firm type, as in this set up I assume a single representative firm. I also assume in this section that elasticities of substitution τ^H, τ^L, σ, as well as age productivity parameters $(\lambda_{H,a})_a$, $(\lambda_{L,a})_a$ do not vary with time. Wages are competitive and equal to marginal productivity:

$$\begin{aligned} w_{H,a}(t) &= \lambda_{H,a}^{\frac{\tau_H - 1}{\tau_H}} H_a(t)^{-\frac{1}{\tau^H}} \times \theta_H(t)^{\frac{\sigma-1}{\sigma}} H(t)^{\frac{1}{\tau_H} - \frac{1}{\sigma}} \times \gamma(t)^{\frac{1}{\sigma}} \quad \forall a \in \{y, m, s\} \\ w_{L,a}(t) &= \lambda_{L,a}^{\frac{\tau_L - 1}{\tau_L}} L_a(t)^{-\frac{1}{\tau^L}} \times \theta_L(t)^{\frac{\sigma-1}{\sigma}} L(t)^{\frac{1}{\tau_L} - \frac{1}{\sigma}} \times \gamma(t)^{\frac{1}{\sigma}} \quad \forall a \in \{y, m, s\} \end{aligned} \tag{22}$$

Which results in relative wage equations:

$$\begin{aligned} \log\left(\frac{w_{H,a}(t)}{w_{H,a'}(t)}\right) &= \frac{\tau_H - 1}{\tau_H} \log\left(\frac{\lambda_{H,a}}{\lambda_{H,a'}}\right) - \frac{1}{\tau^H} \log\left(\frac{H_a(t)}{H_{a'}(t)}\right) \\ \log\left(\frac{w_{L,a}(t)}{w_{L,a'}(t)}\right) &= \frac{\tau_L - 1}{\tau_L} \log\left(\frac{\lambda_{L,a}}{\lambda_{L,a'}}\right) - \frac{1}{\tau^L} \log\left(\frac{L_a(t)}{L_{a'}(t)}\right) \end{aligned} \tag{23}$$

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

$$\begin{aligned} \log\left(\frac{w_{H,a}(t)}{w_{H,a_0}(t)}\right) &= d_{H,a,a_0} - \frac{1}{\tau^H} \log\left(\frac{H_a(t)}{H_{a_0}(t)}\right) + u_{H,a,a_0} \\ \log\left(\frac{w_{L,a}(t)}{w_{L,a_0}(t)}\right) &= d_{L,a,a_0} - \frac{1}{\tau^L} \log\left(\frac{L_a(t)}{L_{a_0}(t)}\right) + u_{L,a,a_0} \end{aligned} \tag{24}$$

Where a_0 is the reference age category. Age productivities $\lambda_{H,a}$, $\lambda_{L,a}$ can then be retrieved from fixed effect d_{H,a,a_0}, d_{L,a,a_0} using normalization conditions (14).

Estimates for aggregate labor inputs $H(t)$ and $L(t)$ can be computed from estimated age productivities and elasticities of substitution. First order conditions (22) also give an expression for relative wage across education levels:

$$\log\left(\frac{w_{H,a}(t)}{w_{L,a}(t)}\right)-\log\left(\frac{\left(\lambda_{H,a}^{\tau_H-1}\frac{H(t)}{H_a(t)}\right)^{\frac{1}{\tau^H}}}{\left(\lambda_{L,a}^{\tau_L-1}\frac{L(t)}{L_a(t)}\right)^{\frac{1}{\tau^L}}}\right)=\frac{\sigma-1}{\sigma}\log\left(\frac{\theta_H(t)}{\theta_L(t)}\right)-\frac{1}{\sigma}\log\left(\frac{H(t)}{L(t)}\right) \tag{25}$$

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

$$\log\left(\frac{w_{H,a}(t)}{w_{L,a}(t)}\right)-\hat{f}=l(t)-\frac{1}{\sigma}\log\left(\frac{H(t)}{L(t)}\right)+v_{a,t} \tag{26}$$

Where $l(t)$ is a linear function of time and $\hat{f}$ is estimated from equations (24).

Weighted Least Square estimation of equations (24) and (26) are presented in table 1 and 2. The weights used are the inverse sampling variance of estimated wage gaps[4]. Labor input from any given education level and age bin is computed as the total sum of hours workers per month in a year. Average wage premia between age and within education are used as outcome variable in equation (24) and computed yearly and by education level by regressing individual wages on a dummy for age, plus fixed effects for gender, industry and region, to control for composition effects. Average wage premia between education levels and within ages are computed in the same fashion.

Table 1: Estimated age productivities and elasticities of substitution - Reduced Form

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

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

[4]In equation (26), I weight by the inverse of the sum of the wage gaps and $\hat{f}$ inverse sampling variance

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

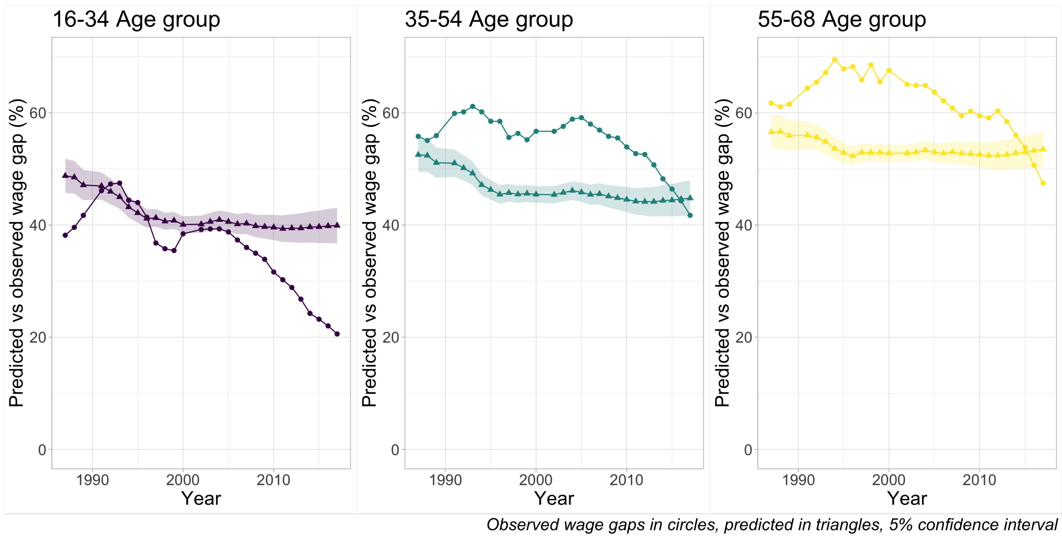
Table 2: Estimated education productivity growth and elasticity of substitution - Reduced Form

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

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

This analysis informs on the large substitutability of workers between age groups and education levels, as well as the slow but significant high school biased technological change occurring in the Portuguese economy between 1987 and 2017, under simple assumptions on supply and demand. Its conclusion is that it is the increase in relative supply of high school graduates that causes the decrease in wage premium, in particular for young workers, who experience a more important rise in relative supply. Due to the high elasticities of substitution however, the effect on wage premia of a rise in relative supply is expected to be modest. Indeed, figure 9 presents the predicted wage gaps by age group. If it matches relatively well observed wage gaps for senior and middle-aged workers, it fails at reproducing the observed drop in young workers' wage premium.

Figure 9: Predicted wage gaps between high school graduates and non-graduates of same age



8 Conclusion

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

References

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

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

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

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. and D. Green
2005. Changes in U.S. Wages, 1976–2000: Ongoing Skill Bias or Major Technological Change? *Journal of Labor Economics*, 23(3):609–648. Publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago].

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.

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

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

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.

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

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

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

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

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.

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

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

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.

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

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

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

A Data

I use *Quadros de Pessoal*, a matched employer-employee dataset provided by the Portuguese National Institute (Instituto Nacional de Estatística, INE). *Quadros de Pessoal* is issued yearly from 1987 to 2017, based on firms declarations on their characteristics and their employees'. Both workers and firms are identified across time by a unique identifier.

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

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

Additionally, I exclude any observation with a missing or 0 worker ID (3.5% of observations per year on average). I am also faced with an issue of duplicate worker IDs which, even though it is minor in the sample later years (about 4.8% of observations per year on average from 2007 to 2017, including 0 IDs), it is much more serious in the earlier years (about 19% of the sample in 1987, including 0 IDs). I suspect these to be encoding mistakes that relate to actual different workers. Some can also be workers who hold two different jobs (for instance an employee somewhere who also have a self-employed activity). Because I do not use the panel aspect of the data, and therefore encoding mistakes in workers ID are not a problem in my analysis, I keep most duplicates, only removing observations who appear more than 5 times in any given year (an average 6.1% of observations per year, less than 1% of the dataset starting in 2007). I also exclude from the sample any worker who is self-unemployed, in unpaid family care, or labelled under "other" employment contract (7.1% of observations per year on average). The rationale behind not considering self-employed is that many of self-employed workers actually work as consultants for a firm, with no way to link them. Self-employed workers on their own represent about 1% of the dataset.

B Proofs

Proposition 1

Proof. Let $Z_1 = \max_{y,k} \{u_{xy}(k)) + \epsilon_{iy}(k)\}$ and $Z_2 = \max_k \{v_y(k) + \eta_j(k)\}$. Then Z_1 follows a Gumbel distribution with expectation $\log \sum_y \int \exp(u_{xy}(k))dk$ and Z_2 follows a Gumbel distribution with expectation $\log \int \exp(v_y(k))dk$

$\mathbb{P}[Z_1 \leq c] = \prod_n \mathbb{P}[u_{xy^n}(k^n) + \epsilon_i^n \leq c]$ is the probability that process $(y^n, k^n, \epsilon_i^n)_n$ has no point in $\{(y, k, \epsilon) : u_{xy}(k) + \epsilon > c\}$. This is:

$$\mathbb{P}[Z_1 \leq c] = \exp\left(-\sum_y \int_{c-u_{xy}(k)} e^{-\epsilon} dk d\epsilon\right)$$

$$\Rightarrow \log \mathbb{P}[Z_1 \leq c] = -\sum_y \int e^{u_{xy}(k)-c} dk = -\exp\left(-c + \log \sum_y \int \exp(u_{xy}(k))dk\right)$$

$\mathbb{P}[Z_2 \leq c] = \prod_n \mathbb{P}\left[v_y(k^n) + \eta_j^n \leq c\right]$ is the probability that process $(k^n, \eta_j^n)_n$ has no point in $\{(k, \eta) : v_y(k) + \eta > c\}$. This is:

$$\mathbb{P}[Z_2 \leq c] = \exp\left(-\int_{c-v_y(k)} e^{-\eta} dk d\eta\right)$$

$$\Rightarrow \log \mathbb{P}[Z_2 \leq c] = -\int e^{v_y(k)-c} dk = -\exp\left(-c + \log \int \exp(v_y(k))dk\right)$$

□

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 $\int_j v_j\, dj$:

$$\begin{aligned} &\inf_{u,v} \int_i u_i\, di + \int_j v_j\, dj \\ &\text{s.t} \int_{i \in k} u_i\, di + v_j \geq \Phi_{y_j}(k) + \int_{i \in k} \epsilon_{iy_j}(k)\, di + \eta_j(k) \quad \forall k, j \\ &\quad u_i \geq \epsilon_{i0} \end{aligned} \tag{27}$$

Take any two u, v such that $\sum_x k_x u_{xy}(k) + v_y(k) \geq \Phi_y(k)$ and $u_{x0} = 0$ and define

$$\begin{cases} u_i = \max_{y,k}\{u_{x_i y}(k) + \epsilon_{iy}(k)\} \\ v_j = \max_k\{v_{y_j}(k) + \eta_j(k)\} \end{cases}$$

Then (u, v) satisfies (27)'s constraints.
Reciprocally, fix any u_i, v_j that satisfy the constraints in this problem and define
Let

$$\begin{cases} u_{xy}(k) = \min_{i,x_i=x}\{u_i - \epsilon_{iy}(k)\} \text{ and } u_{x0} = 0 \\ v_y(k) = \min_{j,y_j=y}\{v_j - \eta_j(k)\} \end{cases}$$

Then the constraint in problem (27) becomes $\sum_x k_x u_{xy}(k) + v_y(k) \geq \Phi_y(k)$.

Applying the law of large numbers, we get that (27) is equivalent to

$$\begin{aligned} &\inf_{u,v} \sum_x n_x G_x(u) + \sum_y m_y H_y(v) \\ &\text{s.t } \sum_x k_x u_{xy}(k) + v_y(k) = \Phi_y(k) \quad \forall k, y \\ &\quad u_{x0} = 0 \end{aligned} \tag{28}$$

By complementary slackness condition, solving problem (27) with $u_{xy}(k) = \alpha_{xy}(k) + w_{xyk}$ and $v_y(k) = \gamma_y(k) - \sum_x k_x w_{xyk}$ yields equilibrium wage. Equilibrium supply and demand $S^x_y(k) = D^y(k)$ obtain as the Lagrange multiplier $\mu_y(k)$ on constraint $\sum_x k_x u_{xy}(k) + v_y(k) \geq \Phi_y(k)$

□

Proof. **Theorem 2**

Rewrite problem (7) as saddle-point:

$$\begin{aligned}
&\inf_{u,v}\sup_{\mu}\sum_x n_x G_x(u)+\sum_y m_y H_y(v)\\
&\quad+\sum_y\int \mu_y(k)\left(\Phi_y(k)-\sum_x k_x u_{xy}(k)-v_y(k)\right)dk+\sum_x S_0^x(-u_{x0})\\
&=\sup_{\mu}\sum_y\int \mu_y(k)\Phi_{ky}\,dk\\
&\quad-\sum_x n_x\sup_u\left\{\sum_y\int \frac{k_x\mu_y(k)}{n_x}u_{xy}(k)+\frac{S_0^x}{n_x}u_{x0}-\sum_x G_x(u)\,dk\right\}\\
&\quad-m_y\sup_v\left\{\int\sum_y \frac{\mu_y(k)}{m_y}v_y(k)-\sum_y H_y(v)\,dk\right\}\\
&=\sup_{\mu}\sum_y\int \mu_y(k)\Phi_{ky}\,dk\\
&\quad-\sum_{x,y}\int k_x\mu_y(k)\log\frac{k_x\mu_y(k)}{n_x}\,dk-\sum_x S_0^x\log\frac{S_0^x}{n_x}-\sum_y\int \mu_y(k)\log\frac{\mu_y(k)}{m_y}\,dk
\end{aligned}$$

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

$$\begin{aligned}
G_x^*(\mu)&=\max_u\left\{\sum_y\int \frac{k_x\mu_y(k)}{n_x}u_{xy}(k)+\frac{S_0^x}{n_x}u_{x0}-\sum_x G_x(u)\,dk\right\}\\
H_y^*(\mu)&=\max_v\left\{\int\sum_y \frac{\mu_y(k)}{m_y}v_y(k)-\sum_y H_y(v)\,dk\right\}
\end{aligned}$$

For which Euler-Lagrange equations are

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

I show that problem (8) attains its maximum in two steps: first, its objective function is continuous in μ under assumption 1 which is shown by applying the dominated convergence theorem. Second, (8) maximizes the social planner function among the set of feasible μ that satisfy:

$$\begin{aligned}
\sum_y\int k_x\mu_y(k)\,dk+S_0^x&=n_x\\
\int \mu_y(k)\,dk&=m_y
\end{aligned}$$

Besides, μ_{yk} is bounded by m_y for every k, so that by the dominated convergence theorem, any sequence $(\mu_n)_n$ that converges weakly converges to a feasible point. Hence the set of feasible μ is compact for the weak topology.

$\square$