Organizations, Skills, and Wage Inequality*

Roberto Pinheiro and Murat Tasci†

January 2017

Abstract

We extend an on-the-job search framework in order to allow firms to hire workers with different skills and skills to interact with firm’s total factor productivity (TFP). Skills not only impact workers' productivity but they also affect how likely workers are able to adapt to changing tasks. Our model implies that more productive firms are larger, they pay higher wages, hire more workers at all skill levels and proportionately more at higher skill types, successfully matching key stylized facts. We calibrate the model using five educational attainment levels as proxies for skills and estimate non-parametrically firm-skill productivity from the wage distributions for different educational levels. We consider two periods in time (1985 and 2009) and three counter-factual economies in which we evaluate how the wage distribution would have evolved if we kept one of the following key characteristics at its 1985’s levels: firm-skill productivity distribution, labor market frictions, and skill distribution. Our results indicate that at least 2/3 of the overall wage dispersion and 1/2 of the within-group component can be attributed to a shift in the productivity distribution conditional on skill. This observed shift over time took the form of first order stochastic dominance for everyone except high-school dropouts. Moreover, changes in labor market frictions, partially counteracted the effect of technology, i.e., between-skill wage dispersion would have been even higher if labor market frictions were kept at its 1985’s levels.

Keywords: Multi-agent firms, skill distributions, wage inequality.

JEL Codes: D02, D21, J2, J3.

*We would like to thank Jim Albrecht, Bruce Fallick, and Susan Vroman for their comments, as well as the participants of the ROA Seminar at Maastricht University, Cleveland State University, and the Cleveland Fed Microeconomics Workshop. Any remaining errors are our responsibility. The views in this article do not necessarily reflect those of the Federal Reserve System or the Board of Governors.

†Federal Reserve Bank of Cleveland, Roberto.Pinheiro@clef.frb.org and Murat.Tasci@clef.frb.org
1 Introduction

Workers with different skills have different career paths. In particular, more educated workers experience lower unemployment rates and lower employment volatility. They also face a higher chance of being offered a better job, being more likely to move on-the-job. For example, Fallick and Fleischman (2001) find that, while the total separation rates fall with worker’s education attainment, in relative terms job-to-job transitions account for a much larger share of total separations for college-educated workers (50%) than for high school drop-outs (30%). Similar evidence has been found by Nagypal (2008), who reports that around 55% of total separations of workers with a college degree are due to job-to-job transitions.

Moreover, highly educated workers are more likely to be employed at large firms, that pay higher wages in general. Studies that look at the labor skill distribution between firms show that larger firms hire over a wider range of skills and as a result have more levels in their hierarchy. Even more, large firms employ proportionally more high-skill workers, i.e., the skill distribution at large firms dominate the skill distribution at smaller firms in first order. Consequently, large firms pay on average higher wages. In fact, large firms pay higher wages than smaller firms for workers with the same observable characteristics. Empirically, these patterns were found by Caliendo, Monte, and Rossi-Hansberg (2015) using data from French manufacturing firms; by Tag (2013) using Swedish data; and Colombo and Delmastro (1999) using Italian data.

However, the data shows that education by itself is not a guarantee of a successful career. Although average wages are increasing with educational attainment, within-group inequality is also increasing in the level of education. As presented by Budría and Telhado-Pereira (2011), education decisions embody wage risks, once recent empirical research has shown that returns to schooling are subject to an important amount of variation across individuals\(^1\). More importantly, these risks are concentrated at the higher levels of education. For example, Lauer (2004) uses data for France and Germany to examine differences across educational qualifications regarding the amount of (unexplained) within-group dispersion. For both countries she finds that wage dispersion is the lowest among workers with a vocational qualification and highest among workers with a tertiary education. Similar results were found by Budría and Telhado-Pereira (2011) for most European countries, among others.\(^2\) Moreover, changes in residual or within-group inequality in the last 30 years has been concentrated at the top-end of the skill distribution. For instance, Lemieux (2006) shows that within-group inequality grew substantially among college-educated workers but changed little for most other groups. Similarly, Autor et. al. \(^1\)

---

\(^1\)Maier et. al. (2004) report that between 20 and 30% of German male workers earn a negative return from schooling while more than 25% earn a return exceeding 15%.

\(^2\)Similar results were found by Buchinsky (1994, 1998) for the United States, Machado and Mata (1997) and Hartog, Pereira, and Vieira (2001) for Portugal, Abadie (1997) for Spain, and Gosling et. al. (2000) for the United Kingdom.
(2005) find that "top-end" residual inequality – e.g. the difference between 90th and 50th percentile of the distribution of residuals – increased substantially while residual inequality at the low-end – the 50-10 gap – actually decreased.

In this paper, we construct a model that allows us to study the market characteristics that generate the patterns for worker’s career paths and organizations’ size and shapes. In our model, not only we allow for firm and worker heterogeneity, but we also allow for firms to hire workers at different skill levels, generating within-firm skill distributions. In this sense, we are able to not only discuss within-group and between-group wage inequality, but also discuss the contribution of within- and between-firm wage dispersion on the overall increase in wage inequality (as discussed by Song et. al. (2016) and Barth et. al. (2016)). Once we calibrate our parameters for two distinct years of the US economy – 1985 and 2009 – we are able to disentangle how much changes in the distribution of firm productivity and educational attainment as well as labor market frictions have contributed to the overall increase in wage inequality. Moreover, we can also evaluate how much those changes may have impacted the within and between-skill group wage inequality.

More specifically, we build a model with frictional labor markets in which firms with different levels of total factor productivity (henceforth, TFP) compete for differently skilled workers that are allowed to search while on-the-job. In our framework, worker’s skill not only affect their productivity, but also how easily they adapt to the changing nature of tasks. New tasks arrive at a Poisson rate \( \delta > 0 \) and a worker is able to solve the new task depending on his/her skill level. We show that in this environment, the patterns found by the empirical literature that connect firm size, firm-skill and wage distributions arise endogenously in equilibrium. We discuss our results in more detail below.

We show that, in equilibrium, firms with higher TFP are larger, employ more workers, and pay higher wages at all skill level. Furthermore, if firms have skill distributions with a shared support, the skill distribution of a firm with higher TFP stochastically dominates in first order the skill distribution of a lower TFP firm. If supports are different, the support of the skill distribution for the low-TFP firms is strictly contained in the support of the skill distributions for the high-TFP firms. If high-skill workers are more productive in home production, the result on stochastic dominance is extended for distributions with different supports.

In terms of wage distributions, we show that earnings are increasing at skill levels, showing that average wages are higher for more skilled workers, independent of the firm’s TFP levels. However, the support of the distributions of wages received by different skill levels may overlap, once workers with different skill levels may receive the same pay at firms with different TFP levels.

In addition, our model allows us to calculate the decomposition of within- and between-firm components of wage dispersion, following previous work by Lazear and Shaw (2007). We also adapt their methodology and decompose the overall wage variance in within- and between-educational group com-
ponents. We present our numerical results in two different ways. First, we show a parametric example based on calibrated parameters for the labor market and education composition, but use parametric estimates for the TFP distribution from previous studies – in particular, Imrohoroglu and Tuzel (2014) – and a Cobb-Douglas production function. We calibrate the production function parameters to match the BLS’s estimates for labor share in 1985 and 2009. This numerical exercise allows us to better understand the model’s key mechanism in addition to giving us some indication that changes in the TFP distribution over time may be the leading reason for the increase in overall wage inequality. However, the examples suffer from either the possibly poor matching of parametric productivity distribution to the model outcomes in labor search models, as pointed by Bontemps et al. (2000), or the use of productivity estimates based on a selected sample (public companies), as well as the need for ad hoc specifications for the production function. In order to avoid these issues, we follow Bontemps et al. (2000) and Launov (2005) and non-parametrically estimate the joint production function. While the methodology does not allow us to separately estimate the TFP distribution and the output distribution, our model provides enough structure in order to allow us to recover the within-firm distribution even without employer-employee matched data nor a clear estimate for the firm’s TFP.

Our numerical results show us that the bulk of the overall wage dispersion (76.6%) between 1985 and 2009 can be explained by differences in productivity across the periods and their interaction with the different skill levels. Changes in the underlying productivity distributions for different skill types account at least half of the within group variation and more than account for the between group variation. Moreover, we see that the productivity (output) distributions for all groups except high school dropouts in 2009 dominate stochastically in first order their 1985 counterparts. This result indicates that skill-biased technological change (SBTC) played a key role for the increase in between-group wage inequality. Our model can successfully match the overall increase in wage dispersion over this period by the observed shift in the productivity distributions as well as the changing educational attainment in the workforce.

We also highlight an increase in dispersion (measured by the standard deviation of earned wages) between 1985 and 2009 that has been concentrated among high educational attainment groups (college grads and post-graduates). This increase in dispersion indicates that firms with different levels of productivity/TFP have quite distinct levels of success leveraging the potential of highly-skilled workers. Consequently, the increase in intra-group wage inequality concentrated at the top of the educational attainment distribution can be explained by some highly skilled workers being lucky to land a job at a firm able to extract a large output from them, while other are unlucky and work for firms in which SBTC was not as intense. As pointed out by Uren and Virag (2011), the fact that luck is a leading cause of within-group inequality may be a feature that helps the model to fit the empirical evidence. The empirical evidence shows that many times the increase in within-group dispersion has been transitory.
(see Gottschalk and Moffitt (1994) and Kambourov and Manovskii (2009), among others), at odds with an explanation based on unobserved skills, since unobserved skills are usually constant or moving slowly over time.\footnote{The literature also considers that changes in the price of unobserved skills due to SBTC are usually persistent, reinforcing the idea that increases in intra-group wage inequality should be long-lasting.}

Finally, our results show that changes in labor market frictions partially undo the effects of the technological changes. Based on our estimates from the CPS MORG, we observe a reduction in labor market frictions at the low educational attainment groups (high-school dropouts and high school graduates) and an increase in frictions for higher educational groups. This feature manifests itself as higher transitions rates into employment and lower separation rates from existing matches for the low skill groups. As our counterfactual results show, average wages and wage dispersion would be even lower for high school dropouts and higher for college graduates and post graduates if labor market frictions were kept at 1985 levels.

Our paper contributes to the literature on skill heterogeneity and wage inequality in search models, by introducing a component of firm-skill distribution. This allows us to eliminate some proposed explanations for the apparent paradox of the consistent increase of between-group inequality in the last 20 years that was not followed by an increase in the within-group inequality. In particular, explanations that would depend on changes in firms’ skill distributions that go against the empirically observed changes in firm hierarchy across firm sizes. In this sense, our model adds to the discussion on within- and between group inequality presented by Lemieux (2006a, 2006b, 2008a, 2008b) and Autor and Acemoglu (2011), as well as the models presented Uren and Virag (2011), and Albrecht and Vroman (2002). We are also able to discuss the different patterns of job mobility between skill levels, as discussed by Dolado, Jansen, and Jimeno (2009), while also addressing and extending the discussion on the relationship between firm size - productivity - and skill distribution presented by Eeckhout and Pinheiro (2014), by allowing within-group inequality, which is not possible in that model. In the same vein, we are able to address the factors that contributed for the increase in between-firm wage inequality, highlighted by Barth et al. (2016) and Song et al. (2016). Finally, our model connects the search literature with the models of knowledge hierarchy (Garicano (2000), and Garicano and Rossi-Hansberg (2006), among others) and Organization adaptability (Dessein and Santos, 2006).

The paper is divided in 4 sections. Section 2 presents the model and our analytical results characterizing the equilibrium. Section 3 discusses our numerical results starting with a parametric exercise that illustrates the main features of the model. Section 3 also presents steady state calibrations for 1985 and 2009 as well as counter-factual exercises that allow us to evaluate how much of the overall increase in within and between-group wage inequality between 1985 and 2009 can be explained by changes in the distribution of educational attainment, changes in labor market frictions, and changes in the joint
distributions of TFP-Skills across the two periods. Finally, section 4 concludes the paper. All proofs are in the appendix.

2 Model

Consider an economy with a measure 1 of firms. Firms have different levels of productivity. We assume a continuous distribution of productivity types $\Gamma(x)$, with support $[\underline{x}, \overline{x}]$. Firms are risk-neutral, infinitely-lived, and discount future at rate $r > 0$. There is a measure $M$ of workers. Workers are heterogeneous in their skills. There are $I$ skills and the measure of workers with skill $i$ is given by $m_i$ with $\sum_{i=1}^{I} m_i = M$. We will initially assume that the skill distribution in the economy is given. All workers are risk-neutral and exit the market at rate $d \in (0, 1)$, being replaced by an unemployed worker of the same skill. Exiting workers experience a disutility cost $c \geq 0$. Workers also discount the future at rate $r > 0$.

The output flow of a worker of skill $i$ in a firm with productivity $x$ is given by $p(x,i)$. We assume that $\frac{\partial p(x,i)}{\partial x} > 0$ and $\frac{\partial p(x,i)}{\partial i} \geq 0$. The output is obtained by successfully completing a given task – failed tasks generate an output flow of zero. We assume that there is a change in tasks faced by a given worker at a Poisson rate $\delta \in (0, 1)$. The new task is drawn from a discrete distribution $H(\cdot)$, with support $\{1, \ldots, J\}$, with $J > N$. In order to simplify exposition, we follow Garicano and Rossi-Hansberg (2006) and assume that the p.d.f. is strictly decreasing, i.e., $h(z) < h(z')$, if $z < z'$. We assume that a worker of skill level $i$ is able to solve tasks in the support $[0,i]$. We consider that there is no team production, in the sense that each worker in the firm either solves a task by herself or not at all. We also assume that workers cannot be reallocated to a different task. Notice that, if the worker cannot solve a problem, given task persistence and a bad enough productivity in this case, the optimal decision by the firm as well as the worker is to destroy the job match. Therefore, a worker of skill level $i$ is dismissed with probability $\delta_i = \delta [1 - H(i)]$, where $H(i) = \Pr(\text{task} \leq i)$. We also assume that a new task can be sampled at the moment that a worker and a firm first match. Consequently, the likelihood that a job match happens at a task in which the worker can successfully execute the job is given by the arrival rate $(1 - \delta_i)$.

The labor market is frictional and search is random. This means that search frictions in the labor market prevent workers from instantaneously matching with the best job offer in the market. Instead, from time to time workers meet randomly with one of the firms in the market. On the other hand, labor markets are partially segmented, so the arrival rate of a job offer may depend on the workers’ skill level. We assume that a worker of skill level $i$ meets a potential employer at rate $\lambda(i) > 0$, irrespective of her employment status. A meeting becomes a job offer at rate $(1 - \Gamma(\overline{\mu}(i)))(1 - \delta_i)$, where $\overline{\mu}(i)$ is the lowest

\footnote{Notice that this is an i.i.d. shock.}
TFP-level firm that hires skill level $i$ (which will be determined in equilibrium). A job offer consists of a wage rate $w$. Worker skills are observable, so firms can condition wage offers to skills. The job offer must be accepted or rejected on the spot, and when rejected it cannot later be recalled. Firms are able to hire everyone who accepts their wage offers, and they choose the wages offered to differently skilled workers in order to maximize their steady state profits. In setting profit maximizing wages, firms take into account both the distribution of wages posted by the other firms in the market. Notice that firms will post one wage for each skill level in the economy. If a firm decides not to hire a given skill level, we will assume the firm is posting a wage rate of 0 for that particular skill.

While searching for a job, unemployed workers engage in home production. We assume that home production by a worker of skill level $i$ produces an output flow rate $b(i)$ with $b'(i) \geq 0$. We also assume that $p(x, 1) > b(1)$ which implies that all firms are active at least at the lowest hung of the skill ladder. We will also keep as a maintained assumption that $b'(i) \geq \frac{\partial p(x, i)}{\partial x}$, i.e., the productivity growth at home production with respect to skill level is at least as large as the increase in productivity at the job at any given type $x$ firm.

2.1 Worker’s Problem

From the framework outlined above, the expected discounted lifetime income when a type $i$ worker is unemployed, $U(i)$, can be expressed as the solution of the following equation:

\[(r + d)U(i) = b(i) - dc + \lambda(i)[1 - \Gamma(x(i))](1 - \delta_i) \int \max\{J(w', i) - U(i), 0\}dF_i(w')\]  

where $b(i)$ is the flow value of home production, $c$ is the cost of exit (injury or death) that arrives at rate $d \in (0, 1)$ and $\lambda(i)[1 - \Gamma(x(i))](1 - \delta_i)$ is the effective arrival rate of a job offer, i.e., the rate at which the type $i$ unemployed workers meet a firm that would like to hire that skill ($\lambda(i)[1 - \Gamma(x(i))])$ and offers a position that the worker can fulfill the required task ($1 - \delta_i$).

Once a type-$i$ worker is employed at a firm paying a wage rate $w$, the value of holding a job at this company is:

\[(r + d)J(w, i) = w - dc + \lambda(i)[1 - \Gamma(x(i))](1 - \delta_i) \int_w (J(w', i) - J(w, i))dF_i(w') + \delta_i(U(i) - J(w, i))\]

where $\delta_i$ is the rate at which a type-$i$ worker faces a change in tasks that make her unproductive at her work. Notice that we already incorporated in the value function the fact that a worker will accept any offer as long as it is above her current wage. This is the case due to the commitment that firms make to a flat wage rate and no counter-offers to outside opportunities an employee may receive.

As $J(w, i)$ is strictly increasing in $w$ whereas $U(i)$ is independent of it, a reservation wage for a worker of skill level $i$, $R(i)$, exists and it is defined by $J(R(i), i) = U(i)$. Then, from equations (1)
and (2), we obtain that:

\[ R(i) = b(i) \]  

(3)

which is driven by the fact that on-the-job search is as effective as out-of-the job search.

Finally, let’s consider the measures of employed and unemployed workers in a steady state equilibrium. Notice that the flows of skill-\( i \) workers in and out of employment in steady state is given by:

\[ d(m_i - u_i) + \delta_i(m_i - u_i) = \lambda(i)(1 - \Gamma(x(i)))(1 - \delta_i)(1 - F_i(R(i)))u_i \]

where the left hand side (henceforth, LHS) of the above equality represents the inflow of skill \( i \) into the unemployment pool, while the right hand side (henceforth, RHS) represents the outflow. Rearranging the above equation, we obtain:

\[ u_i = \frac{(d + \delta_i)m_i}{d + \delta_i + \lambda(i)(1 - \Gamma(x(i)))(1 - \delta_i)(1 - F_i(R(i)))} \]  

(4)

Consequently, the unemployment rate among skill \( i \) workers is given by:

\[ \text{unemp}_i = \frac{u_i}{m_i} = \frac{(d + \delta_i)}{d + \delta_i + \lambda(i)(1 - \Gamma(x(i)))(1 - \delta_i)(1 - F_i(R(i)))} \]  

(5)

In order to present some concepts that will be used in later sections, let’s define \( \gamma_i \) the fraction of skill \( i \) workers among the unemployed population. Given equation (4), we obtain:

\[ \gamma_i = \frac{(d + \delta_i)m_i}{\sum_{j=1}^{I} \frac{(d + \delta_j)m_j}{d + \delta_j + \lambda(j)(1 - \Gamma(x(j)))(1 - \delta_j)(1 - F_j(R(j)))}} \]  

(6)

2.2 Firm’s Problem

In this subsection, we take the behavior of workers as given and derive the firms’ optimal response. Firms post wages that maximize their profits taking as given the distribution of wages posted by their competitors \( (F_i(w), i \in \{1, 2, \ldots, I\}) \) and the distribution of wages that employed workers are currently earning at other firms, given by \( G_i(w), i \in \{1, 2, \ldots, I\} \). We assume here that all distributions are stationary and well-behaved.

When a firm is choosing its optimal wage level at each skill level, it has to take into consideration the amount of workers it can attract at any given wage, as well as if it is optimal to attract a given skill level in the first place. For this reason, before we analyze the firm’s wage decision, let’s derive the distribution of earned wages \( G_i(w), \forall i \in \{1, 2, \ldots, I\} \), as well as the firm’s labor force for each skill level \( i \). In order to do that, let’s start with the firm’s decision of hiring a given skill level or not. In order to hire a type-\( i \) worker, the firm must pay a wage that is at least as high at the reservation wage \( R(i) \). Consequently, a firm with TFP level \( x \) will only hire skill \( i \) if:

\[ p(x, i) \geq R(i) \]
Substituting the expression for $R(i)$ presented in (3), we have:

$$p(x, i) \geq b(i)$$  \hspace{1cm} (7)

Since $p(x, i)$ is strictly increasing on $x$, we have that $x(i)$ – the lowest TFP-level firm that hires skill level $i$ – is determined by the following equality:

$$p(x(i), i) = b(i)$$  \hspace{1cm} (8)

as expected, for any $x > x(i)$, the demand for skill $i$ is strictly positive. Notice that:

$$\frac{\partial p(x, i)}{\partial x} \bigg|_{x=x(i)} \times \frac{dx(i)}{di} + \frac{\partial p(x, i)}{\partial i} \bigg|_{x=x(i)} = b'(i)$$

Rearranging it:

$$\frac{dx(i)}{di} = \frac{b'(i) - \frac{\partial p(x, i)}{\partial i} \bigg|_{x=x(i)}}{\frac{\partial p(x, i)}{\partial x} \bigg|_{x=x(i)}}$$

Since $\frac{\partial p(x, i)}{\partial x} \bigg|_{x=x(i)} \geq 0$, the sign of $\frac{dx(i)}{di}$ depends on the numerator. We will consider 2 cases:

a. $b'(i) = \frac{\partial p(x, i)}{\partial i}$, $\forall i \in I$. This case follows the literature in human capital and search (see Fu (2011) and Burdett, Carrillo-Tudela and Coles (2011)). Consequently, we have $\frac{dx(i)}{di} = 0$ and active firms share a support of skills, i.e., $x(i) = x^*$, $\forall i \in I$;

b. $b'(i) > \frac{\partial p(x, i)}{\partial i}$, $\forall i \in I$. This case considers that autarky productivity increases faster with skills than firm output. We could consider this the entreprenureal route. As expected, here we have $\frac{dx(i)}{di} > 0$.

We now consider the distribution of employed type $i$ workers that earn a wage less than $w$, $G_i(w)$. Let’s initially consider how this distribution evolves over time:

$$\frac{dG_i(w, t)}{dt} = \lambda(i)[1 - \Gamma(x(i))] \left(1 - \delta_i\right) \left[F_i(w, t) - F_i(R(i), t)\right] u_i(t)$$

$$- \left\{d + \delta_i + \lambda(i)[1 - \Gamma(x(i))] \left(1 - \delta_i\right) \left[1 - F_i(w, t)\right]\right\} G_i(w, t) \left(m_i - u_i(t)\right)$$  \hspace{1cm} (9)

The first term on the RHS of (9) describes the inflow at time $t$ of skill $i$ unemployed workers into firms offering a wage no greater than $w$ to skill $i$ workers at time $t$, whereas the second term represents the flow out into death, unemployment, and higher paying jobs, respectively. Since in steady-state $\frac{dG_i(w, t)}{dt} = 0$, this distribution can be rewritten as:

$$G_i(w) = \frac{(d + \delta_i) \left[F_i(w) - F_i(R(i))\right]}{\left\{d + \delta_i + \lambda(i)[1 - \Gamma(x(i))] \left(1 - \delta_i\right) \left[1 - F_i(w)\right]\right\} \left(1 - F_i(R(i))\right)}$$  \hspace{1cm} (10)
Finally, the steady-state number of skill \( i \) workers earning in the interval \([w, w - \varepsilon]\) is represented by 
\[
G(w) - G(w - \varepsilon) \left( m_i - u_i \right)
\]
while 
\[
[1 - \Gamma(x(i))][F(w) - F(w - \varepsilon)]
\]
is the measure of firms offering wages in the same interval. Notice that \([1 - \Gamma(x(i))]\) represents the measure of workers actively hiring skill \( i \) workers. Thus, the measure of skill \( i \) workers per actively hiring \( i \) firms earning a wage \( w \) can be expressed as:

\[
l(w; i) = \frac{(m_i - u_i)}{[1 - \Gamma(x_{\min}(i))]} \lim_{\varepsilon \to 0} G(w | i) - G(w - \varepsilon | i)
\]

Solving it and substituting \( u_i \), we obtain:

\[
l(w; i) = \frac{\lambda(i)(1 - \delta_i)(d + \delta_i)m_i}{\{d + \delta_i + \lambda(i)[1 - \Gamma(x(i))][1 - \delta_i](1 - F_i(w))\}^2}
\]

\[
(11)
\]

### 2.3 The Firm’s Problem and Labor Market Equilibrium

Firms face the problem of picking the wage that maximizes their steady state profits. If a firm pays a higher wage than its peers, ceteris paribus, workers will value being employed in this firm more relative to other firms. As a result, workers employed in other firms are more likely to move into the firm in question whenever they receive this firm’s wage offer. Similarly, when the firm’s own workers receive alternative offers themselves, they are more likely to reject those and stay with the firm. Therefore, with higher wages facilitating recruitment and retention, the firm employs more workers in the steady state. While this force pushes overall profit upwards, it comes at the cost of earning less profit per worker, which pushes profits downwards. At the optimal wage, the firm balances those two counteracting dimensions, maximizing total steady-state profits.

As we mention before, firms have different levels of productivity \( x \). The distribution of productivity levels in the economy is given by a continuous \( \Gamma(\cdot) \), with support \([x, \bar{x}]\). Firms can offer different wages for different skill levels. Therefore, the profit function for a firm that posts wages \( \{w(i)\}_{i=1}^{N} \) and has productivity level \( x \) is given by:

\[
\pi(x) = \sum_{i \in A(x)} (p(x, i) - w(i))l(w(i); i)
\]

where \( A(x) \) denotes the set of skill in which a productivity \( x \) firm is actively searching, i.e., posts a wage above the reservation wage of that skill level. It is easy to see that a firm with productivity \( x \) would post a wage above the reservation wage \( R(j) \) of a given skill \( j \) if \( p(x, j) \geq R(j) \). Therefore, we can define:

\[
A(x) = \{i \in \{1, \ldots, N\} | p(x, i) \geq R(i) \}
\]

Therefore, the firm’s problem is to pin down the wage schedule \( \{w(i)\}_{i \in A(x)} \) in order to maximize \( \pi(x) \). Due to the linearity of the profit function in skills, firms can pin down the wage posted for each
skill level separately, i.e., firm’s can solve:

\[ \pi(x; i) = \max_{w} (p(x, i) - w) l(w; i) \quad (14) \]

for each \( i \in A(x) \).

Then, the optimal wage posted by a firm of productivity \( x \) for a skill level \( i \in A(x) \) is given by \( w = K(x; i) \), given \( p(x, i), F(\cdot | i) \). From the first order condition – i.e., \( \frac{\partial \pi(x, w(i); i)}{\partial w(i)} = 0 \), we obtain:

\[ -l(w; i) + (p(x, i) - w) \frac{dl(w; i)}{dw} = 0 \quad (15) \]

From our expression for \( l_i(w) \), we have that:

\[ \frac{dl_i(w)}{dw} = \frac{2\lambda(i)(1 - \Gamma(\bar{x}(i)))(1 - \delta_i)f_i(w)}{d + \delta_i + \lambda(i)[1 - \Gamma(\bar{x}(i))](1 - \delta_i)(1 - F_i(w))} l_i(w) \quad (16) \]

Substituting it back in the F.O.C., we have:

\[ 2\lambda(i)[1 - \Gamma(\bar{x}(i))](1 - \delta_i)f_i(w)(p(x, i) - w) - \{d + \delta_i + \lambda(i)[1 - \Gamma(\bar{x}(i))](1 - \delta_i)(1 - F_i(w)) \} = 0 \quad (17) \]

which provides us with an implicit equation for \( w \) as a function of \( p(x, i) \) and \( F \). At this stage we proceed as if the second order conditions are satisfied. Below we then verify that this condition is met by the equilibrium solution.

Let’s now consider the optimal wage function \( w(x, i) = K(x, i) \). The flow profit \( \pi(x; i) \) that a skill level \( i \) generates for a firm of productivity level \( x \) offering wage \( K(x; i) \) is:

\[ (p(x, i) - K(x; i)) l(K(x; i); i) \quad (18) \]

Then, a differentiation with respect to \( x \), using the envelope theorem, gives us:

\[ \frac{\partial \pi(x; i)}{\partial x} = \frac{\partial p(x, i)}{\partial x} l(K(x; i); i) \quad (19) \]

Then, we have that:

\[ \pi(x; i) = \int_{x(i)}^{x} \frac{\partial p(x', i)}{\partial x'} l(K(x'; i); i) \, dx' \quad (20) \]

From equation (11) and due to the result that \( [1 - F_i(K(x'; i))] = \frac{[1 - \Gamma(x')]}{[1 - \Gamma(\bar{x}(i))]} \) (see Bontemps et. al. (2002), Proposition 3), we have that:

\[ \pi(x; i) = \int_{x(i)}^{x} \frac{\partial p(x', i)}{\partial x'} \frac{\lambda(i)(1 - \delta_i)(d + \delta_i) m_i}{\{d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(x')]\}^2} dx' \quad (21) \]

Finally, once \( \pi(x; i) = (p(x, i) - K(x; i)) l(K(x; i); i) \Rightarrow K(x; i) = p(x, i) - \frac{\pi(x; i)}{l(K(x; i); i)} \). Then, substituting \( \pi(x; i) \), we have:

\[ K(x; i) = p(x, i) - \int_{x(i)}^{x} \frac{\partial p(x', i)}{\partial x'} \left( \frac{d + \delta_i + \lambda_1 (1 - \delta_i) [1 - \Gamma(x)]}{d + \delta_i + \lambda (1 - \delta_i) [1 - \Gamma(x')]} \right)^2 dx' \quad (22) \]

Before we move on, let’s present some important basic results.
Lemma 1 For any skill level, the wage is increasing in firm productivity, i.e., $\frac{\partial K(x,i)}{\partial x} > 0$.

Moreover, define $l_i(x) \equiv l_i(K(x,i))$. Then, we get the additional results:

Lemma 2 $\frac{\partial l_i(x)}{\partial x} \geq 0$, $\forall i \in I$. Consequently, firms with higher TFP hire more of each skill.

Then, define firm size of a firm with TFP $x$ by:

$$S(x) = \sum_{i \in I(x)} l_i(x)$$

(23)

Then, a simple corollary of lemma 2 follows (keeping in mind that $\frac{dx(i)}{dt} \geq 0$).

Corollary 1 Firm size is increasing in $x$, i.e., $S(x) > S(x')$ if $x > x'$.

Now, let’s define the skill distribution at a $x$-type firm by:

$$\Phi_x(i) = \sum_{i'=1}^{i} \frac{l_{i'}(x)}{S(x)}$$

(24)

We then want to show the following proposition:

Proposition 1 If $x > x'$ we have that $\Phi_x(i) \leq \Phi_{x'}(i)$, $\forall i \in I$. Consequently $\Phi_x(i)$ F.O.S.D. $\Phi_{x'}(i)$.

The next theorem collects all the results we presented up to now.

Theorem 1 In this economy, we are able to show that firms with a higher total factor productivity, in equilibrium:

- Are bigger;
- Hire more at all skill levels;;
- Hire proportionately more at the top;;
- Hire all the skill that firms with lower TFPs hire and may also hire workers with higher skills that their lower TFP counterparts – i.e. the support of skills hired by high TFP firms strictly contain the support of skills hired by low TFP firms;;
- Pay higher wages at all skill levels.

Before we move on to the description of the distributions of posted and earned wages of different skill types, let’s present the average wage and standard deviation of wages within-firms, as a function of the firm TFP. As we will see in later sections, these expressions will be important in our discussion of the decomposition of wage dispersion in within and between-firms components. Based on our previous
calculations, we can easily show that the average wage and variance of the within-firm wage distribution as a function of the firm’s TFP $x$ is given by:

$$E_{\phi_x}[w] = \frac{\sum_{i=1}^{\ell} K(x,i)l_x(i)}{S(x)}$$  \hspace{1cm} (25)$$

and

$$Var_{\phi_x}[w] = \frac{\sum_{i=1}^{\ell} (K(x,i) - E_{\phi_x}[w])^2l_x(i)}{S(x)}$$  \hspace{1cm} (26)$$

respectively.

### 2.3.1 Wage distributions by skill type

In this subsection, we will look how the equilibrium distributions of posted and earned wages vary across worker types. In this sense, we will be able to evaluate how workers with different abilities are faced with different job market opportunities and outcomes. In order to do that, we will initially focus on how firms’ wage posting strategy vary with workers’ skills. In particular, based on the optimal wage posting strategy presented in equation (22), we obtain the following result.

**Lemma 3** \(\frac{\partial K(x,i)}{\partial i} > 0\) i.e., wages at any TFP level increase with skill level.

Therefore firms with higher productivity offer higher wages at all skill levels, and workers with higher skill levels receive higher wage offers from all different productivity levels. Notice that this does not mean that there is no overlap between the distributions of offered wages for different skills, once the wage offered by firms with different productivity levels to workers with different skill levels may be the same. In fact, as a corollary of the Lemma 3, we know that if two firms offer the same wage to workers of different skill levels, the firm that offer this particular wage to the higher skilled worker must have a lower total factor productivity.

**Corollary 2** If a given wage $w$ is offered for both skills $i$ and $j$, $i > j$, we must have that the firm offering wage $w$ for skill level $i$ has lower productivity than the firm offering the wage for skill $j$, i.e., $w = K(x,i) = K(x',j) \Rightarrow x < x'$.

Similarly, gathering our previous results and our assumptions on $b(\cdot)$ and $p(\cdot, \cdot)$ and taking into account the traditional Burdett and Mortensen (1998) arguments (so distribution are continuous with connected supports), we also have the following corollary:

**Corollary 3** If $i > j$, we have that the support of offered wages are given by $[b(i), K(\bar{x}, i)]$ and $[b(j), K(\bar{x}, j)]$, where $b(i) \geq b(j)$ and $K(\bar{x}, i) > K(\bar{x}, j)$. 

13
Then, considering the distribution of posted wages by firms and how the firms’ total factor productivity affects its posted wages, we can show the following result.

**Proposition 2** If \( i > j \), \( F_i(w) \) FOSD \( F_j(w) \).

As a straightforward consequence of FOSD, we have

**Corollary 4** The average offered wage increases with skill, i.e., if \( i > j \), \( E_{F_i}[w] \geq E_{F_j}[w] \).

Once we obtained these results for the distribution of posted wages, we are now able to present some key characteristics for the distributions of earned wages by skill levels. From our previous calculations, we have:

\[
G_i(w) = \frac{(d + \delta_i)F_i(w)}{(d + \delta_i + \lambda(i)(1 - \delta_i)(1 - \Gamma(K^{-1}(w_i, i))))(1 - \delta_i)(1 - F_i(w))}
\]

while the p.d.f. is given by:

\[
g_i(w) = \begin{cases} 
\frac{(d + \delta_i)(d + \delta_i + \lambda(i)(1 - \delta_i))(1 - \Gamma(K^{-1}(w_i, i)))}{(d + \delta_i)(1 - \Gamma(K^{-1}(w_i, i)))(1 - F_i(w))} f_i(w) & \text{if } w \in [w_i, w_i] \\
0 & \text{otherwise}
\end{cases}
\]

We are now able to show the following proposition:

**Proposition 3** If \( i > j \), \( G_i(w) \) FOSD \( G_j(w) \).

Then, a simple corollary of the previous result is presented next.

**Corollary 5** The average earned wage increases with skill, i.e., if \( i > j \), \( E_{G_i}[w] \geq E_{G_j}[w] \).

### 2.3.2 Economy-wide wage distributions

In this subsection, we present the economy-wide distributions of offered and earned wages. Our goal is to show how the wage distributions per skill type interact in order to build their economy-wide counterpart and consequently, how changes in the composition of the labor force may affect the economy-wide wage distribution. We start with the distribution of posted wages, followed by the distribution of earned wages.

**Aggregated density of posted wages**

First of all, let’s be very precise about the densities for each skill level \( i \):

\[
f_i(\bar{w}) = \begin{cases} 
\frac{\gamma(K^{-1}(\bar{w}, i))}{[1 - \Gamma(K^{-1}(\bar{w}, i)) \frac{\partial K^{-1}(\bar{w}, i)}{\partial K^{-1}(\bar{w}, i)}}] & \text{if } \bar{w} \in [w_i, w_i] \\
0 & \text{otherwise}
\end{cases}
\]
Then, notice that due to the fact that not all firms offer jobs at all skill levels, we need to weight the wage distributions per skill by the measure of firms that actively post wages at that particular skill level. Once weights are properly included, the aggregated cumulative distribution of offered wages in the economy is given by:

\[ F(w) = \frac{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))] F_i(w)}{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))]} \tag{30} \]

Consequently, the density function associated to this cumulative distribution is given by:

\[ f(w) = \frac{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))] f_i(w)}{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))]} \tag{31} \]

Similarly, the average posted wage for the overall economy is given by:

\[ E_F[w] = \int_w^{\bar{w}} \bar{w} f(w) dw = \frac{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))] \int_{w_i}^{\bar{w}} \bar{w} f_i(\bar{w}) d\bar{w}}{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))]} \tag{32} \]

Simplifying the expression and substituting the average offered wage per skill, we have:

\[ E_F[w] = \frac{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))] E_F[w]}{\sum_{i=1}^{I} [1 - \Gamma(K^{-1}(w, i))]} \tag{33} \]

**Aggregated density of earned wages**

In this case, we focus on the wage that workers are actually earning. Consequently, instead of tracking the measure of firms posting a job at a given skill level, we need to track the measure of employed workers at each skill level as a proportion of all employed workers. Based on the cumulative distribution of earned wages per skill by the measure of firms that actively post wages at that particular skill level, we have that the aggregated cumulative distribution of earned wages is given by:

\[ G(\bar{w}) = \frac{\sum_{i=1}^{I} G_i(\bar{w})(m_i - u_i)}{\sum_{i=1}^{I} (m_i - u_i)} \tag{34} \]

while the corresponding density function is given by:

\[ g(\bar{w}) = \frac{\sum_{i=1}^{I} g_i(\bar{w})(m_i - u_i)}{\sum_{i=1}^{I} (m_i - u_i)} \tag{35} \]

Based on these results, the average aggregate wage in this economy is given by:

\[ E_G[w] = \int_w^{\bar{w}} w g(w) dw = \int_w^{\bar{w}} \frac{\sum_{i=1}^{I} g_i(w)(m_i - u_i)}{\sum_{i=1}^{I} (m_i - u_i)} dw = \frac{\sum_{i=1}^{I} (m_i - u_i) E_G_i[w]}{\sum_{i=1}^{I} (m_i - u_i)} \tag{36} \]

where \( E_G_i[w] \) can be derived after some manipulations and a change in variables considering \( w = K(x, i) \):

\[ E_G_i[w] = \int_{\bar{g}(i)}^{x'} K(x', i) (d + \delta_i) \times \left[ \frac{d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(\bar{g}(i))] - \gamma(x')}{(d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(\bar{g}(i))])^2} \right] \frac{\gamma(x')}{[1 - \Gamma(\bar{g}(i))]} dx' \tag{37} \]
Finally, let’s consider the economy-wide variance of earned wages:

$$Var_G(w) = \int \frac{w}{\bar{w}} (w - E_G(w))^2 g(w)dw$$

(38)

Again, after some algebra and a change of variables considering $w = K(x,i)$, as well as defining:

$$M - U = \sum_{i=1}^{I}(m_i - u_i)$$

we have:

$$Var_G(w) = \frac{1}{M - U} \int \sum_{i=1}^{I}(K(x,i) - E_G(w))^2 \frac{\lambda(i)(1 - \delta_i)(d + \delta_i)m_i}{\{d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(x)]\}^2} \gamma(x)dx$$

(39)

Substituting equation (11), we can rewrite $Var_G(w)$ as:

$$Var_G(w) = \frac{1}{M - U} \int \sum_{i=1}^{I}(K(x,i) - E_G(w))^2 l_x(i)\gamma(x)dx$$

(40)

### 2.4 Wage Variance Decompositions

As we described in the introduction, one of the most important questions that our model can address pertains to the source of wage inequality across workers in the overall economy. In particular, the decomposition of the source of wage dispersion in terms of the within versus between firms components is one that has been the focus of most of the recent empirical literature. In particular, according to Lazear and Shaw (2008), the total variance in wages, $\sigma^2$ is given by the following:

$$\sigma^2 = \sum_{j=1}^{J} p_j \sigma_j^2 + \sum_{j=1}^{J} p_j (\bar{w}_j - \bar{w})^2$$

(41)

The first term on RHS of equation (41) is the within-firm component of the variance. Notice that $p_j$ is the share of workers in the economy employed in firm $j$, while $\sigma_j^2$ is the variance of wages in firm $j$. The second term on the RHS of equation (41) represents the between-firms component of the wage variance. In this expression, $\bar{w}_j$ is the mean wage in firm $j$, and $\bar{w}$ is the mean wage in the economy.

In this section, we apply their decomposition to our model. Let’s start with the within-firm component. Assume that we partition the interval $[\underline{x}, \overline{x}]$ in $N$ intervals of length $\Delta$. Then, we have $x_{i+1} = x_i + \Delta$. Moreover, the measure of type $x_{i+1}$ firms is given by $\Gamma(x_{i+1}) - \Gamma(x_i)$, while the share of employed workers in the economy at type $x_{i+1}$ is $\frac{S(x_{i+1})}{M - U}$. Then, we have that $p_j \sigma_j^2$ can be rewritten as $Var_{\phi_{x_{i+1}}}(w) \frac{S(x_{i+1})}{M - U} \Gamma(x_{i+1}) - \Gamma(x_i)$. Multiplying and dividing by $\Delta$ and adding across $x_S$, we have:

$$\sum_{i=1}^{N} Var_{\phi_{x_{i+1}}}(w) \frac{S(x_{i+1})}{M - U} \frac{[\Gamma(x_{i+1}) - \Gamma(x_i)]}{\Delta}$$

Taking $\Delta \to 0$, we have:

$$\text{Within-firm component} = \int \frac{S(x)}{M - U} Var_{\phi_x}(w)\gamma(x)dx$$

(42)
Following the same procedure for the between-firms component, we obtain:

\[
\text{Between-firms component} = \int_{x}^{\pi} \frac{S(x)}{M - U} (E_{\phi_x}(w) - E_G(w))^2 \gamma(x) dx
\]  \hfill (43)

Therefore, the entire decomposition can be rewritten as:

\[
Var_G(w) = \int_{x}^{\pi} \frac{S(x)}{M - U} Var_{\phi_x}(w) \gamma(x) dx + \int_{x}^{\pi} \frac{S(x)}{M - U} (E_{\phi_x}(w) - E_G(w))^2 \gamma(x) dx
\]  \hfill (44)

where \( Var_{\phi_x}(w), E_{\phi_x}(w), \) and \( E_G(w) \) are presented in equations (26), (25), and (36), respectively.

Moreover, we can adapt Lazear and Shaw (2008)’s decomposition in order to decompose the overall wage variance in terms of the within and between educational groups. Using the same logic, we obtain the following decomposition:

\[
\sigma^2 = \sum_{i=1}^{I} m^e_i (\sigma^e_i)^2 + \sum_{i=1}^{I} m^e_i (\overline{w}^e_i - \overline{w})^2
\]  \hfill (45)

Similar to the decomposition in terms of within and between firms, the first term in the RHS of equation (45) represents the within-skill component of the overall wage inequality, and the second term in the RHS is the between-skills component. \( m^e_i \) is the fraction of the employed labor force with skill level \( i \), \( \sigma^e_i \) is the standard deviation of wages for workers of skill \( i \), \( \overline{w}^e_i \) is the average wage for employed workers with skill level \( i \), and \( \overline{w} \) is the mean wage in the economy.

Mapping this decomposition to our model, we have \( \overline{w}^e_i = E_{G_i}[w] \), where \( E_{G_i}[w] \) is given by equation (37), while \( (\sigma^e_i)^2 \) is given by:

\[
(\sigma^e_i)^2 \equiv Var_{G_i}(w) = \frac{(d+\delta_i)\{d+\delta_i+\lambda(i)(1-\delta_i)[1-\Gamma(\xi(i))])\}}{\lambda(i)(1-\delta_i)[1-\Gamma(\xi(i))]} \times \int_{\xi(i)}^{\pi} \frac{K(x, i) - E_{G_i}[w])^2}{[d+\delta_i+\lambda(i)(1-\delta_i)[1-\Gamma(x(i))]} dx
\]  \hfill (46)

Finally, the fraction of employed workers that have skill level \( i \) in our model is given by:

\[
m^e_i = \frac{\lambda_i(1-\delta_i)[1-\Gamma(\xi(i))]m_i}{d+\delta_i+\lambda_i(1-\delta_i)[1-\Gamma(\xi(i))]} \sum_{j=1}^{I} \frac{\lambda_j(1-\delta_j)[1-\Gamma(\xi(j))]m_j}{d+\delta_j+\lambda_j(1-\delta_j)[1-\Gamma(\xi(j))]} \]

\hfill (47)

Consequently, the decomposition of overall wage variance in terms of within- and between-skills is given by:

\[
\text{Within-skill component} = \sum_{i=1}^{I} m^e_i Var_{G_i}(w) \quad \text{and} \quad \text{Between-skills component} = \sum_{i=1}^{I} m^e_i (E_{G_i}[w] - E_G[w])^2
\]  \hfill (48)

where \( m^e_i, Var_{G_i}(w), E_{G_i}[w], \) and \( E_G[w] \) are given by equations (47), (46), (37), and (36), respectively. In the next sections we will use both within- and between-firms and within- and between-skills decompositions in order to define the sources of the increase in the wage dispersion between 1985 and 2009.
3 Quantitative Results

In this section, we gaze the performance of the model with some quantitative results focusing on its performance over time. Our model incorporates two sided heterogeneity into a search framework and has important implications about the relationship between worker skills and firm productivities. We assume that the worker heterogeneity in terms of skill is well measured by educational attainment. This allows us to calibrate our model in terms of skill endowments directly from the data. However, calibrating firm-level heterogeneity in terms of productivity is not as straightforward. Estimating a firm-level TFP distribution from establishment level data is beyond the scope of this paper. Instead, we proceed in two different ways. First, we take a publicly available estimate of the distribution of $x$ and impose a particular functional form on firm-level match output. This allows us to evaluate the quantitative implications of the model. Our second set of results come from an exercise where we are agnostic about the actual functional form of the match-output. We instead non-parametrically generate the implied productivity distributions from the observed wage distributions conditional on skill levels.

3.1 Parametric Example

While most of the parameters in the exercise follow the calibrated parameters presented in the next section, we deviate from the full calibration exercise in two ways. First, in this section, we consider a production function per match that follows a Cobb-Douglas, i.e., the match output is given by:

\[ p(x, i) = x^{\alpha} i^{\beta}. \] (49)

Second, we use the estimates for total factor productivity for public companies in the U.S. in order to obtain a distribution of TFP at two different time periods: 1985 and 2009. The estimates are from Imrohoroglu and Tuzel (2014), which are obtained by implementing the methodology of Olley and Pakes. We then apply a non-parametric Kernel estimation in order to pin down the distribution. Figure 1 presents the two density functions. As we can observe, the density for 2009 is close to a mean-spread preserve of the 1985’s distribution. The 1985 distribution has a mean 0.7789 and standard deviation 0.3070, while the 2009 distribution has a mean 0.7725 and standard deviation 0.4372. This result is in line with what Faggio et al. (2010) observed for the U.K.

There are benefits and drawbacks in this approach. The main benefit is that by explicitly considering a production function and using the standard approach to obtain the TFP, we are able to disentangle $x$, $i$, and $p(x, i)$. Consequently, we are able to not only separate the effects of TFP and education on the output flow, but also single out a firm as well as a firm’s TFP, allowing us to discuss within and between-firms wage dispersion. On the other hand, not only we must impose a particular (and stable) functional form for production, but we also need to rely on estimates for TFP from a selected sample of firms, which is biased towards large firms.
Table 1 presents the calibrated parameters for labor market frictions, educational attainment, as well as unemployment benefits and/or home production income for both 1985 and 2009. Section 3 describes how we calibrate the parameters for the educational attainment distribution and the labor market frictions. The parameters for \( b(i) \) are specific for the parametric calibration and are pinned down such that the lowest TFP firm makes zero profits by hiring any of the skills. Moreover, we pin down \( \alpha \) and \( \beta \) such that the \( \beta \) is equal to the Bureau of Labor Statistics (BLS) estimate for the labor share in years 1985 and 2009, respectively, and that \( \alpha + \beta = 1 \) in both years. Consequently, we have that, in 1985 \( \alpha = 0.386 \) and \( \beta = 0.614 \) and in 2009 \( \alpha = 0.417 \) and \( \beta = 0.583 \).

Table 1: Parameters

<table>
<thead>
<tr>
<th></th>
<th>A: 1985</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( i )</td>
<td>( m_i )</td>
<td>( \lambda(i) )</td>
<td>( \delta_i )</td>
<td>( b(i) )</td>
<td></td>
</tr>
<tr>
<td>( i = 1 )</td>
<td>0.1843</td>
<td>0.4782</td>
<td>0.0584</td>
<td>0.2220</td>
<td></td>
</tr>
<tr>
<td>( i = 2 )</td>
<td>0.4039</td>
<td>0.4538</td>
<td>0.0287</td>
<td>0.3398</td>
<td></td>
</tr>
<tr>
<td>( i = 3 )</td>
<td>0.1998</td>
<td>0.5459</td>
<td>0.0233</td>
<td>0.4359</td>
<td></td>
</tr>
<tr>
<td>( i = 4 )</td>
<td>0.1234</td>
<td>0.5290</td>
<td>0.0125</td>
<td>0.5201</td>
<td></td>
</tr>
<tr>
<td>( i = 5 )</td>
<td>0.0886</td>
<td>0.5236</td>
<td>0.0093</td>
<td>0.5965</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B: 2009</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( i )</td>
<td>( m_i )</td>
<td>( \lambda(i) )</td>
<td>( \delta_i )</td>
<td>( b(i) )</td>
<td></td>
</tr>
<tr>
<td>( i = 1 )</td>
<td>0.0936</td>
<td>0.5464</td>
<td>0.0530</td>
<td>0.1513</td>
<td></td>
</tr>
<tr>
<td>( i = 2 )</td>
<td>0.2665</td>
<td>0.4870</td>
<td>0.0274</td>
<td>0.2267</td>
<td></td>
</tr>
<tr>
<td>( i = 3 )</td>
<td>0.2917</td>
<td>0.5187</td>
<td>0.0227</td>
<td>0.2871</td>
<td></td>
</tr>
<tr>
<td>( i = 4 )</td>
<td>0.2258</td>
<td>0.5031</td>
<td>0.0123</td>
<td>0.3396</td>
<td></td>
</tr>
<tr>
<td>( i = 5 )</td>
<td>0.1223</td>
<td>0.4998</td>
<td>0.0097</td>
<td>0.3868</td>
<td></td>
</tr>
</tbody>
</table>

As we can observe in table 1, while labor frictions seemed to decrease for low education groups – both high school dropouts and high school graduates observed an increase in \( \lambda_i \) and a decrease in \( \delta_i \) between 1985 and 2009 – all other groups experienced an increase in labor market frictions.

Figure 2 shows how average wages and wage dispersion vary with skill and TFP for 1985. As expected, given Lemmas 1 and 3, average wages increase with skill and TFP. However, graphs (c) and (d) in figure 2 show an interesting result that we are not able to predict based on the analytical results. FOSD in figure does not allow us to make predictions about standard deviation. As we can see, wage dispersion increases with TFP as well as educational attainment, i.e., intra-firm wage dispersion increases with TFP and intra-group wage dispersion goes up with education. These results are in line with the findings in the empirical literature.

Figure 3 compares the results for 1985 presented in figure 2 against the results in 2009. As we can see average wages increased, although significantly less than what has been observed in the data. This
is a drawback in the parametric calibration, likely due to our choice of the output function. Notice
that the average earned wage increased by 5.69 percent for high school graduates, while it increased by
4.18 percent for college grads and 4.04 percent for post-graduates. Moreover, the standard deviation of
wages increased by 24.31 percent for high school dropouts, while increasing by 37.10 and 37.28 percent
for college grads and post-grads, respectively. Finally, as we see in graph (c) in figure 3, the standard
deviation of within-firm variance also increased for high-TFP firms. In order to further evaluate the
changes in within and between-firms wage inequality, we consider the decomposition proposed by Lazear
and Shaw (2008). Results are presented in table 2, panel A. As we can see in table 2, within-firm wage
dispersion actually decreased, while between-firms wage dispersion increased significantly in the period,
pushing the overall wage variance up by 14.53 percent. Similarly, we can also present a decomposition
of the overall wage dispersion in terms of within and between-skill groups, as discussed in previous
sections. Results are presented in the panel B of table 2. As we can observe, the overall increase in
wage dispersion is mostly due to an increase in within-group wage dispersion.

<table>
<thead>
<tr>
<th>Table 2: Variance Decomposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Firms</strong></td>
</tr>
<tr>
<td>Var(_G(w))</td>
</tr>
<tr>
<td>Within-firm</td>
</tr>
<tr>
<td>Between-firms</td>
</tr>
<tr>
<td>1985 0.4440</td>
</tr>
<tr>
<td>2009 0.5085</td>
</tr>
<tr>
<td><strong>B: Skills</strong></td>
</tr>
<tr>
<td>Var(_G(w))</td>
</tr>
<tr>
<td>Within-group</td>
</tr>
<tr>
<td>Between-group</td>
</tr>
<tr>
<td>1985 0.4440</td>
</tr>
<tr>
<td>2009 0.5085</td>
</tr>
</tbody>
</table>

3.2 Non-parametric Calibration

As we discussed before, while the results presented based on our parametric exercise have the advantage
of allowing us to discuss within and between-firms wage inequality, it comes at the cost of depending on
the TFP estimates based on a biased sample and the imposition of a particular functional form for the
production function. In this section, we consider a non-parametric calibration of the output flow \(p(x, i)\)
following Bontemps et. al. (2000). This methodology allows us to use the wage distribution in order
to recover the productivity distribution, circumventing the need for firm-level data. As a drawback, we
are unable to separately estimate \(x\) and \(p(x, i)\). Consequently, our estimates are not directly related to
a given firm. However, we are still able to discuss within- and between-groups wage dispersion.

We calibrate the steady-state model for the U.S. economy at two different periods (1985 and 2009).
We use these different calibrations in order to evaluate the contribution of changes in the productivity distribution, the educational attainment distribution, and labor market frictions to the increase in overall wage inequality, as well as to the change in within and between-skill groups components.

In order to calibrate our parameters, we use wage and educational data from the 1985 Current Population Survey’s Merged Outgoing Rotation Groups (CPS MORG) as well as the 2009 American Community Survey (ACS). From the 1985 CPS MORG, we classify 5 different educational attainment levels based on Jaeger (1997)’s method, allowing us to compare the results with the classification presented at the 2009 ACS. We decided to work with the 2009 ACS instead of the 2009 CPS MORG due to the higher values for top-coding in earnings presented at the ACS. We also use data from the CPS MORG’s of 1985 and 2009 in order to obtain data for the average employment to unemployment (EU) transition rates, as well as the average transition from unemployment to employment (UE) for each education group.

We calibrate the parameters in our model considering 1 month as the unit of time. We calibrate the death rate \( d \) in order to match the average overall death probability rate at the median age of 37 throughout the entire period (1985-2009). Moreover, for each time period (1985 and 2009) we calibrate the job finding rate \( \lambda(i)(1 - \delta_i)(1 - \Gamma(x(i))) \) to match the average UE transition for each education group from the CPS data. Similarly, we calibrate the job destruction rate \( \delta_i \) such that \( \delta_i + d \) match the average EU transition rate in the data for each education group. Finally, we follow Bontemps et al. (2000) and Launov (2005) to estimate \( p(x, i) \). In particular, we follow the steps presented below:

1. We non-parametrically (kernel estimation) estimate \( G_i(w) \). We use Launov’s procedure to obtain an estimate of the upper tail using a Pareto distribution;
2. We adjust the estimates in order to obtain distributions that are the closest to the estimated ones but still satisfy the models conditions: Single-peaked and

\[
3\kappa_i g_i(w)^2 - g_i'(w) [1 + \kappa_i G_i(w)] > 0 \tag{50}
\]

where \( \kappa_i = \frac{\lambda(i)(1-\Gamma(x(i))(1-\delta_i))}{d+\delta_i} \)
3. We assume all firms hire all skill levels (\( \Gamma(x(i)) = 0, \forall i \)).

Based on our estimates for the earned wage distributions and the parameters \( \lambda(i), \delta(i), \) and \( d \), estimated from the job flows and death rates, we pin down productivity and productivity distribution as:

\[
p(x, i) = w + \frac{1 + \kappa_i G_i(w)}{2\kappa_i g_i(w)} \tag{51}
\]

\[
\gamma(p(x, i)) = \frac{2\kappa_i (1 + \kappa_i) g_i(w)^3}{3\kappa_i \{1 + \kappa_i G_i(w)\}^2 g_i(w)^2 - g_i'(w) \{1 + \kappa_i G_i(w)\}^3} \tag{52}
\]
Note that the recovered productivity distributions will be consistent with the labor market turnover rates and the average unemployment rates by different skill groups. Between conditions 2 and 3 described above, 2 ends up being the most material. It effectively guarantees that wages are non-decreasing with the firm-productivity for a given skill-type. This does not stand out as a very restrictive assumption, a priori, but we find this to be violated in the data, especially at the low-skill levels and at the low-end of the wage distribution among those worker types. Single-peak feature guarantees that for wages higher than the mod-wage, this condition is satisfied just by the virtue of $g'_i(w) < 0$. We ensure this by effectively flattening the local peaks in the relevant domain for each skill-types, keeping the aggregate mass constant. This step is rather straightforward and does not require a calibrated wage-density that is far-off from the empirical one. The problem is a bit complicated on the left side of the distribution for wages lower than the mod-wage. Intuitively, our restriction implies that the calibrated density cannot be increasing ‘very fast’ in that region. It turns out that this seems to be a feature of the data anyway, for higher skill types. In the end, we can easily calibrate productivity distributions so that we can get the wage densities exactly for workers with a college degree or more, both for 1985 and 2009. For the lower-skilled workers though, our best fit seems to fail to account for the rapid increase in the density early on.

We present the empirical estimates of $g_i(w)$ and the closest distributions we can get following our methodology that satisfies conditions 2.) and 3.) in figures 4 and 5 for 1985 and 2009, respectively. As we can see, the adjusted distributions, that serve as the inputs in our non-parametric calibration of the output distributions, are quite close to the empirical counterparts. Our largest ‘mismatch’ is for the high-school dropouts in 1985, and that yields a fitted wage density that has 5.6 percent of its mass away from the empirical density. For the rest of the skill types it rapidly declines to 3 percent for high-school graduates and 1 percent for workers with some college. By these measures, our fit is much better for 2009. We obtain a calibrated density that only relocates 4.5 percent of the mass for lowest skill type and 1.5 percent for the high school grads. The mismatch for workers with some college education is less than 0.5 percent. For both years, our non-parametric match to the empirical density for college graduates and workers with post-graduate degrees requires a distortion that is less than 0.03 percent of the respective mass. In summary, our inputs to the calibration of the output distributions present only small deviations from the empirical counterparts, allowing us to obtain a good fit even though we have a quite parsimonious model.

3.2.1 Benchmark results: 1985 and 2009

We now present our results for the calibrations for 1985 and 2009. First of all, as we can see in figure 6, there is clear evidence of FOSD for high skill levels, in particular for skill levels 4 (college graduates) and 5 (post graduates). This skill bias technological change (SBTC) can explain the significant increase
in wages for high skill levels. Differently, for lower levels of educational attainment we see either minor movements (for high school graduates and some college) or a technical change that may have even hindered productivity – for high school dropouts. Moreover, as we can see in figure 7, while the FOSD can be clearly seen in panel (a), with the high increases in average output for college graduates and post-graduates, panel (b) shows that this increase was accompanied by a large increase in within-group output inequality for high skill workers. In fact, while college graduates and post-graduates have seen an increase in the standard deviation of output flow of 26.32 and 26.45 percent between 1985 and 2009, high school graduates and workers with some college have only seen an increase in output standard deviation of 0.74 and 9.79 percentage points. Even more, high school dropouts have actually seen a drop in output dispersion of -27.35 percent.

In terms of changes in the wages per worker as a function of educational attainment, results are as expected. The model shows that, while average wages have increased significantly for college graduates and post-graduates – by 20 and 24.06 percent, respectively – it moved just by 2.27 percentage points for high-school graduates and 7.93 percent for workers with some college, and it went down by 7.75 percent for high school dropouts. In terms of the within-group standard deviation, we see that wage dispersion has significantly increased for high-skill workers, while moving only slightly at the middle of the educational attainment distribution and has even gone down among high school dropouts.

<table>
<thead>
<tr>
<th>Table 3: Variance Decomposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skills</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Var(w)</td>
</tr>
<tr>
<td>1985   2432229.61</td>
</tr>
<tr>
<td>2009   4393235.46</td>
</tr>
<tr>
<td>Within-group</td>
</tr>
<tr>
<td>1985   1920437.77</td>
</tr>
<tr>
<td>2009   3250064.73</td>
</tr>
<tr>
<td>Between-group</td>
</tr>
<tr>
<td>1985   511791.85</td>
</tr>
<tr>
<td>2009   1143170.74</td>
</tr>
</tbody>
</table>

As we can see in table 3, while the overall variance increased by 80.63 percent, about two thirds of the increase (67.80 percent) was due to an increase in within-group variance, while only about a third (32.20 percent) happened due to an increase in between-groups variance. These results reinforce the importance of the evidence presented in figures 6 and 7, i.e., that SBTC jointly with an increase in output dispersion, in particular for high-skill workers, drove up within-group wage dispersion and consequently, overall wage dispersion.

Finally, in terms of how much of the overall variation observed in the data our model can capture, table 4 compares the average wage and within-group standard deviation for the model and data in 1985 as well as 2009, everything expressed in 2009 dollars. In order to avoid a situation where the inputed tail dominates the overall standard deviation, we cap wages at five times the top-coded wages for both 1985 and 2009 both in the data and in the inputs used to estimate the output distributions. As we see, results show that the model does a good job capturing the within-group variation across groups.
Moreover, results do not seem to fundamentally depend on the imputation method used to obtain the censored upper tail. In an on line appendix, we follow Dustmann, Ludsteck, and Schonberg (2007) and use different methods to impute censored wages. All methods deliver results that are qualitatively the same and similar in magnitude. As we can see in table 4, the model does a decent job capturing the wage variation in both time periods and across the educational groups, in particular at the lower educational groups. In fact, the model tends to underestimate both the average wage and within-skill wage dispersion for college graduates and post-graduates.

Table 4: Comparison Model vs. Data

A. Average Earned Wage

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>HS Dropout</th>
<th>HS Graduate</th>
<th>Some College</th>
<th>College</th>
<th>Post-Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985 - Model</td>
<td>3187.24</td>
<td>2379.43</td>
<td>2952.60</td>
<td>3175.56</td>
<td>3961.3</td>
<td>4676.21</td>
</tr>
<tr>
<td>1985 - Data</td>
<td>3271.34</td>
<td>2412.44</td>
<td>2869.40</td>
<td>3267.33</td>
<td>4152.78</td>
<td>5006.63</td>
</tr>
<tr>
<td>2009 - Model</td>
<td>3813.66</td>
<td>2194.93</td>
<td>3019.57</td>
<td>3427.46</td>
<td>4753.40</td>
<td>5801.39</td>
</tr>
<tr>
<td>2009 - Data</td>
<td>4018.60</td>
<td>2252.78</td>
<td>3062.28</td>
<td>3483.42</td>
<td>5099.25</td>
<td>6423.04</td>
</tr>
</tbody>
</table>

B. St. Deviation of earned wages

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>HS Dropout</th>
<th>HS Graduate</th>
<th>Some College</th>
<th>College</th>
<th>Post-Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985 - Model</td>
<td>1559.56</td>
<td>1098.04</td>
<td>1351.69</td>
<td>1400.00</td>
<td>1583.39</td>
<td>1663.36</td>
</tr>
<tr>
<td>1985 - Data</td>
<td>1720.91</td>
<td>1147.35</td>
<td>1402.67</td>
<td>1608.26</td>
<td>1899.38</td>
<td>1987.5</td>
</tr>
<tr>
<td>2009 - Model</td>
<td>2096.00</td>
<td>980.58</td>
<td>1488.88</td>
<td>1689.05</td>
<td>2177.95</td>
<td>2285.43</td>
</tr>
<tr>
<td>2009 - Data</td>
<td>2471.72</td>
<td>1002.48</td>
<td>1500.39</td>
<td>1760.06</td>
<td>2761.96</td>
<td>3087.53</td>
</tr>
</tbody>
</table>

3.2.2 Counterfactual experiments

In this section, we consider a few counterfactual exercises. In principle, our model has three sets of parameters/inputs: labor market frictions – defined by $\lambda_i, \delta_i, d, i \in \{1, 2, ..., 5\}$ – educational attainment distributions, characterized by $m_i, i \in \{1, 2, ..., 5\}$, and output per worker-skill distributions, obtained through the non-parametric estimation using earned wage data. In our counterfactual exercises, we will calculate how much of the increase in wage inequality – both overall as well as within- and between-skill groups – can be explained by changes in each one of the set of inputs separately. In order to do that, we start from the 1985 benchmark calibration inputs and consider that only one of the three sets of inputs (labor market frictions, educational distribution, and output distributions) is changed to its 2009 counterpart. We then compare the obtained wage distribution against the 1985 and 2009 benchmark results. Moreover, in some cases we also consider the additional contribution of moving another input towards its 2009 values in terms of how much closer we get to the 2009 wage distributions.
Table 5 as well as figure 9 summarize the results. Table 5 shows the standard deviation of earned wages for each educational group in 1985 as well as in 2009 and their difference. Then, it shows how much of this difference can be explained by each counterfactual. Notice that the production counterfactual, that keeps labor market frictions and the educational attainment distribution at their 1985 levels, while moving the production distribution to their 2009 levels, explains more than 100 percent of the difference for all but high school graduates. This result shows that the technological change not only explains the bulk of the result, but also that changes in the labor market frictions helped to attenuate the impact of SBTC. We would also like to highlight that once the change in standard deviation of earned wages was negative for high school dropouts between 1985 and 2009, a percentage above 100 percent in the production scenario implies that we should have expected an even bigger drop in wage dispersion for high school dropouts in this scenario. Panel (b) in figure 9 gives a visual perspective for the result. Moreover, while the scenario in which we only update the educational attainment distribution seems to not affect the within-group dispersion, it actually affects the overall dispersion and the break down between within- and between-groups, as we can see in table 6.

Table 6 shows not only the difference in overall variance across the different time periods and counterfactuals, but also how the overall wage inequality is decomposed in terms of within and between groups. As in table 5, the percentages in the counterfactual rows show how much of the difference between the 1985 and 2009 values is observed in the counterfactual scenario. As we can see, 76.6 percent of the overall increase in wage dispersion can be explained by just the change in productivity distribution. Moreover, as we see, there is a clear interaction between education and productivity parameters that boost between group wage dispersion, while labor market frictions again acts in the opposite direction.

Finally, table 7 shows the impact of the changes in the output distribution, educational attainment, and labor market frictions between 1985 and 2009 on the average wages earned by different skill groups,
Table 6: Variance Decomposition - Skills

<table>
<thead>
<tr>
<th></th>
<th>Var_G(w)</th>
<th>Within-group</th>
<th>Between-group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>2432229.61</td>
<td>1920437.77</td>
<td>511791.85</td>
</tr>
<tr>
<td>2009</td>
<td>4393235.46</td>
<td>3250064.73</td>
<td>1143170.74</td>
</tr>
<tr>
<td>Difference</td>
<td>1961005.85</td>
<td>1329626.96</td>
<td>631378.89</td>
</tr>
</tbody>
</table>

Counterfactuals:
- Production: 76.76% 52.54% 127.78%
- Labor: 2.35% 11.98% -17.93%
- Education: 5.78% 17.10% -18.05%
- Education + Production: 112.33% 52.54% 238.26%

as well as the average wage in the overall economy. Results are in line with what we found for wage dispersion. Changes in technology clearly boosted the wages for high skill workers, while helped to depress the average wage of low skill workers. Differently, changes in the labor market frictions seem to partially undo the effect of the technological change. While changes in the education distribution are not relevant for the average wage of particular educational groups, it is quite relevant to explain the increase in overall average wages, in particular once combined with the technological changes.

Table 7: Average Earned Wages

<table>
<thead>
<tr>
<th></th>
<th>HS Dropout</th>
<th>HS Grad</th>
<th>Some College</th>
<th>College</th>
<th>Post Graduate</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Benchmark</td>
<td>2194.93</td>
<td>3019.57</td>
<td>3427.46</td>
<td>4753.4</td>
<td>5801.39</td>
<td>3813.66</td>
</tr>
<tr>
<td>1985 Benchmark</td>
<td>2379.43</td>
<td>2952.6</td>
<td>3175.56</td>
<td>3961.3</td>
<td>4676.21</td>
<td>3187.24</td>
</tr>
<tr>
<td>Difference</td>
<td>-184.5</td>
<td>66.97</td>
<td>251.9</td>
<td>792.1</td>
<td>1125.18</td>
<td>626.42</td>
</tr>
</tbody>
</table>

Counterfactuals:
- Production: 195.93% -115.17% 114.95% 110.01% 115.06% 31.37%
- Labor: -131.83% 243.48% -12.43% -7.11% -10.53% 36.94%
- Education: 0.00% 0.00% 0.00% 0.00% 0.00% 36.94%
- Education + Labor: -131.83% 243.48% -12.43% -7.11% -10.53% 40.81%
- Production + Labor: 100.00% 100.00% 100.00% 100.00% 100.00% 39.08%
- Education + Production: 195.93% -115.17% 114.95% 110.01% 115.06% 100.47%

4 Conclusion

In this paper, we present a model that can decompose the overall wage dispersion into dispersion between and within skill groups, as well as within and between firms, while delivering most of the properties discussed in the empirical literature in organizations as equilibrium properties. We calibrate
the model both parametrically and non-parametrically using wage data from the CPS MORG and ACS for the years 1985 and 2009. Our results show that technological change induced both an increase in the average wage and wage dispersion for highly educated workers, while depressing average wages and reducing dispersion among low-education workers. Changes in labor market frictions helped to partially undo the effect of technology. Changes in the educational attainment, while not affecting the average wage and wage dispersion within particular educational groups, helped to boost overall wage dispersion and average wages. In terms of the components of wage dispersion, we see that the increase in overall wage dispersion was mostly due to an increase in the within-group component.

Overall, our results suggest that SBTC was a major factor for the increase in wage inequality between educational groups. Our results for the parametric calibration using TFP estimates for public companies suggest that most of the increase in wage inequality is due to the between-firm component. However, this result is not conclusive due to the bias in the sample of firms used for the estimation (large public firms).
REFERENCES


Maier, Michael, Friedhelm Pfeiffer, and Winfried Pohlmeier, “Returns to Education and Individual Heterogeneity,” 2004, mimeo.


Papageorgiou, Theodore “Large Firms and Within Firm Occupational Reallocation”, mimeo.

Song, Jae, David Price, Fatih Guneven, Nicholas Bloom, and Till von Wachter, “Firming up Inequality”, 2016, mimeo.


Appendix

Proof of Lemma 1

Proof. Taking the derivative with respect to \( x \), we have:

\[
\frac{\partial K(x, i)}{\partial x} = 2[d + \delta_i + \lambda(i)(1 - \delta_i)](1 - \Gamma(x))\lambda(i)(1 - \delta_i)\gamma(x) \times \\
\times \int_0^x \frac{\partial p(x', i)}{\partial x'} \left( \frac{1}{d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(x)]} \right)^2 \, dx'
\]

Consequently, \( \frac{\partial K(x, i)}{\partial x} > 0. \) ■

Proof of Lemma 2

Proof. Taking the derivative with respect to \( x \), we get:

\[
\frac{\partial l_i(x)}{\partial x} = \frac{2\lambda(i)(1 - \delta_i)(d + \delta_i)m_i\{d + \delta_i + \lambda(i)(1 - \delta_i)(1 - \Gamma(x))\}}{\{d + \delta_i + \lambda(i)(1 - \delta_i)(1 - \Gamma(x))\}^3} > 0.
\]

■

Proof of Proposition 1

Proof. Based on the results obtained up to now, we know that the support of skills for a type \( x \) firm is \( \{1, ..., I(x)\} \) while the support of skills hired by firm \( x' \) is \( \{1, ..., I(x')\} \) with \( I(x') \leq I(x) \). Consequently, if \( I(x') \leq i \leq I(x) \) we have that \( \Phi_{x'}(i) = 1 \) and \( \Phi_x(i) \leq 1 \) and consequently \( \Phi_x(i) \leq \Phi_{x'}(i) \). Let’s then now consider the case in which \( i < I(x') \). In this case, both firms hire this particular skill. Since the distributions are discrete, let’s focus on the p.d.f.s \( \phi_x(i) \) and \( \phi_{x'}(i) \). Then, notice that

\[
\phi_x(i) = \frac{l_x(i)}{S(x)}
\]

As we showed before:

\[
\frac{d l_i(x)}{d x} = \frac{2\lambda(i)(1 - \delta_i)\gamma(x)}{d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(x)]}l_i(x)
\]

and

\[
S'(x) = \sum_{j=1}^{I} \frac{2\lambda(j)(1 - \delta_j)\gamma(x)}{d + \delta_j + \lambda(j)(1 - \delta_j)[1 - \Gamma(x)]}l_j(x)
\]

Then:

\[
\frac{\partial \phi_x(i)}{\partial x} = \frac{1}{S(x)^2} \sum_{j=1}^{I} 2\gamma(x)l_i(x)l_j(x) \left[ \frac{\lambda(i)(1 - \delta_i)}{d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(x)]} - \frac{\lambda(j)(1 - \delta_j)}{d + \delta_j + \lambda(j)(1 - \delta_j)[1 - \Gamma(x)]} \right]
\]

Rearranging it, we have:

\[
\frac{\partial \phi_x(i)}{\partial x} = \frac{1}{S(x)^2} \sum_{j=1}^{I} 2\gamma(x)l_i(x)l_j(x) \left\{ \frac{d[\lambda(i)(1 - \delta_i) - \lambda(j)(1 - \delta_j)] + \delta_j\lambda(i)(1 - \delta_i) - \delta_i\lambda(j)(1 - \delta_j)}{[d + \delta_i + \lambda(i)(1 - \delta_i)[1 - \Gamma(x)] \times [d + \delta_j + \lambda(j)(1 - \delta_j)[1 - \Gamma(x)]]} \right\}
\]

31
So the sign of the derivative depends on the numerator of the terms within curly brackets. Keep in mind that if \( i > j \) we have that \( \delta_i < \delta_j \) and we are also keeping the assumption that \( \lambda(i) \geq \lambda(j) \). Consequently, we have that \( \lambda(i)(1 - \delta_i) > \lambda(j)(1 - \delta_j) \) and \( \delta_j \lambda(j)(1 - \delta_j) > \delta_i \lambda(j)(1 - \delta_j) \). So, the terms in the summation in which \( i < j \) are negative, while the terms in which \( i > j \) are positive. As expected if \( i = j \) the term is zero.

Based on these results, we have that \( \frac{\partial \phi_x(1)}{\partial x} < 0 \) and \( \frac{\partial \phi_x(f)}{\partial x} \geq 0 \) (with inequality if \( x > x(I) \)). Now let’s consider the C.D.F. Since the variable is discrete, the C.D.F. is a step function. We therefore just need to see how the steps move up or down with \( x \). Since we showed that \( \frac{\partial \phi_x(1)}{\partial x} < 0 \), we know that the first step moves down with \( x \). Let’s then consider the second step:

\[
\frac{\partial \phi_x(1)}{\partial x} + \frac{\partial \phi_x(2)}{\partial x} = \frac{1}{S(x)^2} \sum_{i=1}^{I} \sum_{j=3}^{I} 2\gamma(x)l_i(x)l_j(x) \left\{ \begin{array}{l}
\frac{d[\lambda(i)(1 - \delta_i) - \lambda(j)(1 - \delta_j)]}{\delta_i - \delta_j} + \\
\frac{+\delta_j \lambda(i)(1 - \delta_i) - \delta_i \lambda(j)(1 - \delta_j)}{\delta_i - \delta_j} \\
[\delta_i + \lambda(i)(1 - \delta_i)(1 - \Gamma(x))] \times \\
\times [\delta_j + \lambda(j)(1 - \delta_j)](1 - \Gamma(x))
\end{array} \right. \]

where, since the positive term in the summation for \( i = 2 \) cancels out with a negative term in \( i = 1 \), we only have negative terms left and \( \frac{\partial \phi_x(1)}{\partial x} + \frac{\partial \phi_x(2)}{\partial x} < 0 \). Consequently, the second step also goes down with \( x \). The same argument can be made for the next step. In fact, considering a given \( i' < I \), we have that:

\[
\sum_{i=1}^{I} \frac{\partial \phi_x(i)}{\partial x} = \frac{1}{S(x)^2} \sum_{i=1}^{I} \sum_{j=i'+1}^{I} 2\gamma(x)l_i(x)l_j(x) \left\{ \begin{array}{l}
\frac{d[\lambda(i)(1 - \delta_i) - \lambda(j)(1 - \delta_j)]}{\delta_i - \delta_j} + \\
\frac{+\delta_j \lambda(i)(1 - \delta_i) - \delta_i \lambda(j)(1 - \delta_j)}{\delta_i - \delta_j} \\
[\delta_i + \lambda(i)(1 - \delta_i)(1 - \Gamma(x))] \times \\
\times [\delta_j + \lambda(j)(1 - \delta_j)](1 - \Gamma(x))
\end{array} \right. < 0
\]

Consequently, increasing \( x \) pushes the steps down at any \( i \). Therefore, we have that \( \Phi_x(i) \leq \Phi_x(i'), \forall i \in \mathcal{I} \), implying that \( \Phi_x(i) \) F.O.S.D. \( \Phi_x(i') \).

**Proof of Lemma 3**

**Proof.**

\[
\frac{\partial K(x,i)}{\partial x} = \frac{\partial p(x,i)}{\partial x} - \int_{x(i)}^{x} \frac{\partial^2 p(x',i)}{\partial x'^2} \left( \frac{d+\delta_i + \lambda(i)(1-\delta_i)\Delta}{d+\delta_i + \lambda(i)(1-\delta_i)\Delta}[1-\Gamma(x')] \right) \frac{d\Gamma(x')}{dx'} dx' \\
+ 2 \int_{x(i)}^{x} \frac{\partial p(x',i)}{\partial x'} \left( \frac{d+\delta_i + \lambda(i)(1-\delta_i)\Delta}{d+\delta_i + \lambda(i)(1-\delta_i)\Delta}[1-\Gamma(x')] \right) \frac{d\Gamma(x')}{dx'} dx' \\
+ \frac{\partial \phi_x(i)}{\partial x} \left( \frac{d+\delta_i + \lambda(i)(1-\delta_i)\Delta}{d+\delta_i + \lambda(i)(1-\delta_i)\Delta}[1-\Gamma(x')] \right)^2 \frac{d\phi_x(i)}{dx}.
\]

Notice that:
From Corollary 2, we have that \( \Gamma(w, i) > \Gamma(w, j) \) if \( x > x' \). Moreover, from Corollary 3, we have that the support of wages offered to \( i \) and \( j \)-type workers are \([b(i), K(\pi, i)]\) and \([b(j), K(\pi, j)]\), where \( b(i) \geq b(j) \) and \( K(\pi, i) > K(\pi, j) \). Moreover, the distributions are continuous with no mass points and with a connected support, as showed by Burdett and Mortensen (1998). Now let’s consider \( F_i(w) \) and \( F_j(w) \). If \( w \in (b(j), b(i)) \), we have that \( F_j(w) > 0 \) and \( F_i(w) = 0 \), so \( F_i(w) \leq F_j(w) \). Similarly, if \( w \in (K(\pi, j), K(\pi, i)) \) we have that \( F_j(w) = 1 \) and \( F_i(w) < 1 \), so \( F_i(w) \leq F_j(w) \). Finally, if \( w \in [b(i), K(\pi, j)] \), so the wage is offered to both type \( i \) and \( j \) workers. From previous calculations, we have that:

\[
F_i(w) = \frac{\Gamma(K^{-1}(w, i)) - \Gamma(K^{-1}(w, i))}{1 - \Gamma(K^{-1}(w, i))}
\]

From Corollary 2, we have that \( K^{-1}(w, i) \) \( < K^{-1}(w, j) \), since the firm offering the same wage \( w \) for a higher skill worker must have a TFP than the firm offering the same wage for a lower skilled worker. Consequently, \( \Gamma(K^{-1}(w, i)) < \Gamma(K^{-1}(w, j)) \). On the other side, since based on our assumptions we have \( \frac{dx}{dt} \geq 0 \), we have that \( \bar{x}(t) \equiv K^{-1}(w, i) \geq K^{-1}(w, j) \) \( \equiv \bar{x}(j) \). Consequently \( \Gamma(K^{-1}(w, i)) \geq \Gamma(K^{-1}(w, j)) \). Then, we have that:

\[
F_i(w) - F_j(w) = \frac{\Gamma(K^{-1}(w, i)) - \Gamma(K^{-1}(w, j))}{1 - \Gamma(K^{-1}(w, i))} - \frac{\Gamma(K^{-1}(w, j)) - \Gamma(K^{-1}(w, j))}{1 - \Gamma(K^{-1}(w, j))}
\]
Manipulating it, we have:

\[
F_i(w) - F_j(w) = \begin{cases} 
\frac{\Gamma(K^{-1}(w, j)) [1 - \Gamma(K^{-1}(w, i))]}{[1 - \Gamma(K^{-1}(w, j))]} - \frac{\Gamma(K^{-1}(w, i)) [1 - \Gamma(K^{-1}(w, j))]}{[1 - \Gamma(K^{-1}(w, i))]} \\
+\frac{\Gamma(K^{-1}(w, i)) - \Gamma(K^{-1}(w, j))}{[1 - \Gamma(K^{-1}(w, j))]} \\
\end{cases} < 0
\]

Adding and subtracting one to the numerator and manipulating, we obtain:

\[
F_i(w) - F_j(w) = \begin{cases} 
\frac{[1 - \Gamma(K^{-1}(w, i))] [1 - \Gamma(K^{-1}(w, j))] - [1 - \Gamma(K^{-1}(w, j))] [1 - \Gamma(K^{-1}(w, i))]}{[1 - \Gamma(K^{-1}(w, j))] [1 - \Gamma(K^{-1}(w, i))]} \\
\end{cases} < 0
\]

Since \([1 - \Gamma(K^{-1}(w, i))] < [1 - \Gamma(K^{-1}(w, j))]\) and \([1 - \Gamma(K^{-1}(w, j))] < [1 - \Gamma(K^{-1}(w, i))]\). Consequently \(F_i(w) \leq F_j(w), \forall w\). Therefore \(F_i(w) \text{ FOSD } F_j(w)\). ■

**Proof of Corollary 4**

**Proof.**

\[
E_{F_i}[w] = \int_{\mathbb{R}} K(x', i) \frac{\gamma(x')}{1 - \Gamma(z(i))} dx' \geq \int_{\mathbb{R}} K(x', j) \frac{\gamma(x')}{1 - \Gamma(z(j))} dx' \geq \int_{\mathbb{R}} K(x', j) \frac{\gamma(x')}{1 - \Gamma(z(j))} dx' = E_{F_j}[w]
\]

Where the first inequality comes from the fact that \(\frac{\partial K(x, i)}{\partial x} > 0\), while the second inequality comes from \(F_i(w) \text{ FOSD } F_j(w)\). Now simplifying the above expression, we have \(E_{F_i}[w] \geq E_{F_j}[w]\). ■

**Proof of Proposition 3**

**Proof.** From Lemmas 1 and 3, we know that for all \(x, K(x, i) > K(x, j)\) and for all skill \(i\), we have \(K(x, i) > K(x', i)\) if \(x > x'\). Moreover, from Corollary 3, we have that the support of wages offered to \(i\) and \(j\)-type workers are \([b(i), K(x, i)]\) and \([b(j), K(x, j)]\), where \(b(i) \geq b(j)\) and \(K(x, i) > K(x, j)\).

Moreover, the distributions are continuous with no mass points and with a connected support, as showed by Burdett and Mortensen (1998). Now let’s consider \(G_i(w)\) and \(G_j(w)\). If \(w \in (b(j), b(i))\), we have that \(G_j(w) > 0\) and \(G_i(w) = 0\), so \(G_i(w) \leq G_j(w)\). Similarly, if \(w \in (K(x, j), K(x, i))\) we have that \(G_j(w) = 1\) and \(G_i(w) < 1\), so \(G_i(w) \leq G_j(w)\). Finally, if \(w \in [b(i), K(x, j)]\), so the wage is earned to both type \(i\) and \(j\) workers. In this case, we have:

\[
G_i(w) - G_j(w) = \begin{cases} 
\frac{(d+\delta_i)F_i(w)}{(d+\delta_j)F_i(w)} \frac{1 - \Gamma(K^{-1}(w, i))[1 - \Gamma(K^{-1}(w, j))]}{1 - \Gamma(K^{-1}(w, i))[1 - \Gamma(K^{-1}(w, j))]} \\
- \frac{(d+\delta_j)F_j(w)}{(d+\delta_j)F_j(w)} \frac{1 - \Gamma(K^{-1}(w, j))[1 - \Gamma(K^{-1}(w, j))]}{1 - \Gamma(K^{-1}(w, j))[1 - \Gamma(K^{-1}(w, j))]} \\
\end{cases}
\]

From our assumptions, we know that, since \(i > j\), we have \(\delta_i < \delta_j\) and \(\lambda(i) > \lambda(j)\). Moreover, from previous calculations, we have that:

\[
[1 - \Gamma(K^{-1}(w, i))] [1 - F_i(w)] = [1 - \Gamma(K^{-1}(w, i))]
\]

34
But then, based on Lemmas 1 and 3, we have \[1 - \Gamma(K^{-1}(w, i)) > 1 - \Gamma(K^{-1}(w, j))\]. Finally, from proposition 2, we have \(F_i(w) < F_j(w)\). Putting all these results together, we have:

\[(d + \delta_i)F_i(w) < (d + \delta_j)F_j(w)\]

and

\[\{d + \delta_i + \lambda(i)[1 - \Gamma(K^{-1}(w, i))](1 - \delta_i)[1 - F_i(w)]\} > \{d + \delta_j + \lambda(j)[1 - \Gamma(K^{-1}(w, j))](1 - \delta_j)[1 - F_j(w)]\}\]

and consequently \(G_i(w) - G_j(w) < 0\). In summary, \(G_i(w) \leq G_j(w), \forall w\). Therefore \(G_i(w)\) FOSD \(G_j(w)\). ■
Figure 1: TFP Distributions based on COMPUSTAT Data: 1985 vs. 2009

0 0.5 1 1.5 2 2.5 3 3.5 4
\gamma(x)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
1985
2009
Figure 2: Average and Standard Deviation of wages as a function of Educational Attainment and TFP in 1985
Figure 3: Average and Standard Deviation of wages as a function of Educational Attainment and TFP: 1985 vs. 2009
Figure 4: Comparison between empirical wage distributions and adjusted distributions that fulfill model requirements – 1985
Figure 5: Comparison between empirical wage distributions and adjusted distributions that fulfill model requirements – 2009
Figure 6: Comparison between Cumulative Distributions of outputs for different skill levels – 1985 vs. 2009
Figure 7: Average and Standard Deviation of output for different skill levels – 1985 vs. 2009

Figure 8: Average and Standard Deviation of wages for different skill levels – 1985 vs. 2009
Figure 9: Average and Standard Deviation of wages - Counterfactuals