The Role of Headhunters in Wage Inequality: It's All about Matching[*]

Alexey Gorn[‡]

October 9, 2020[§]

Abstract

This study relates the increase in the U.S. top wages to the increasing prominence of headhunters. Headhunters improve the matching between firms and employees via two channels: screening of candidates and passive on-the-job search. I incorporate headhunters in the labor market framework of random search with two-sided heterogeneity. The calibrated model shows that headhunters can account for 32% of the increase in the top 10% wage share in the U.S. from 1970 to 2010, with 19% due to improvements in matching between workers and firms. I provide supporting micro evidence for CEO compensation, as well as cross-country evidence on headhunter hires/fees and top income growth.

JEL: E24, D83, C78, J24, J62, J63

Keywords: wage distribution, top incomes, sorting, on-the-job search, headhunters

[†]I am grateful to the Editor, Vincenzo Quadrini, and two anonymous referees whose comments and suggestions greatly improved the paper. Many thanks to Antonella Trigari, Nicola Gennaioli, Nicola Pavoni, and Yuri Tserlukevich for valuable discussions and advice. Also thanks to Hamish Low, Nicolas Serrano-Velarde, Philipp Kircher, Tito Boeri, Vasco Carvalho and seminar participants at Bocconi PhD Research Seminar with Alberto Alesina, 2014 EDGE Jamboree in Copenhagen, Workshop on Dynamic Macroeconomics in Vigo, EEA Congress in Geneva, ES EWM in Edinburgh, University of Vienna, IIES, University of Chile, CEU, University of Alicante, University of Melbourne, HSE, NES and University of Cambridge. I gratefully acknowledge financing by Fondazione CARIPLO Thesis Research Grant. Declarations of interest: none.

[‡]University of Liverpool Management School, Chatham Street, Liverpool, L69 7ZH, United Kingdom, e-mail: a.gorn@liverpool.ac.uk.

[§]For published version, visit https://doi.org/10.1016/j.red.2020.10.006

1 Introduction

Wages at the top of the distribution have been rising sharply in the United States since the early 1970s. Top 10% wage share[1] increased from 25.7% in 1970 to 34.5% in 2010 in the U.S. (Piketty (2014)). One of the reasons for the rise of top wages and wage dispersion is improved matching between firms and employees in the top positions[2] (Song et al. (2019)). Why did the matching improve? The conventional view attributes the improvement in matching to skill-biased technological change which raises incentives for firms and workers to be better matched. However, while this explains the rise of the upper-middle class, it cannot explain the sharp rise of top wages. Song et al. (2019) show a strong non-linearity in the sorting pattern[3]: a disproportional shift of high-skilled workers to high-paying firms in comparison to medium-skilled workers to medium-paying firms, and this cannot be generated by technological progress in models of labor markets.

In this paper, I show that the improved matching at the top is due to decreasing search frictions in the labor market for top positions. I develop a model where frictions are reduced by the increasing role of headhunters, or executive search firms. Headhunters started to gain market share in the U.S. in the 1970s and now assist in filling more than half of the positions in the top wage segment. They enhance matching for two reasons. First, they provide more suitable candidates for the firm because they can screen the candidates better. Second, they induce passive on-the-job search as they contact potential candidates directly, creating opportunities for new matches without active search from employed workers[4]. As a result, the headhunters restrict the pool of potential candidates to only the high-skilled workers while, at the same time, expand the pool of potential candidates to a larger number of those high-skilled workers. These two features result in better matching between high-skilled workers and firms, a higher surplus and, therefore, a higher wage in such matches. Because headhunters operate mostly on the top wage segment, such improvements in matching do not happen (or happen to a lower degree) over the rest of the distribution. Therefore, the presence of headhunters generates a strong non-linearity in sorting improvement that leads, in turn, to soaring of top wages compared to the rest of the distribution. The decrease in the search frictions also changes the outside options of workers and firms affected by the headhunters. The changes in the outside options affect the wage bargaining outcome and increase the top wages even

[1]The share of total wages that goes to the top 10% of all employees.

[2]Alternative explanations of increasing wage inequality include: i) decrease in top income taxes - Alvaredo et al. (2013); ii) direct effects of skill-biased technological change on wages - Acemoglu (2002); Autor, Katz and Kearney (2006); and many others; iii) social norms - Piketty (2014); iv) exogenous changes in random growth theories - Jones (2015); Gabaix et al. (2016); Aoki and Nirei (2017); Jones and Kim (2018); v) numerous studies on the increase of CEO pay including Gabaix and Landier (2008); Lemieux, MacLeod and Parent (2009); and Bell and Reenen (2013) among others.

[3]Bagger, Sørensen and Vejlin (2013) document similar findings using Danish data. They find that the correlation between worker and firm fixed effects increased from -0.07 in 1981 to 0.14 in 2001. For the top quartile of workers the correlation increased from -0.20 to 0.37, while the correlation stayed almost unchanged at around zero for the rest of the quartiles.

[4]Faberman et al. (2017) show that a significant share of unsolicited contacts and offers go to employed workers not looking for another job.

further at the same time compressing the wages at the bottom.

To quantify the contribution of better matching induced by headhunters to the increase in top wages, I develop a labor market model along the lines of Pissarides (1985) augmented with heterogeneous workers and firms[5]. I introduce the headhunter industry by adding a new channel for matching workers and firms. Firms with an open position can either post a vacancy as in the standard model or hire through a headhunter. The difference for the firm is that it cannot screen workers coming through vacancies, while the headhunter guarantees a minimal skill level of the worker with whom the firm is matched. Consider now the worker's side. Low-skilled workers have access to the standard channel and they can search from both unemployment and employment. Every worker searching through the standard channel has to pay a per-period search cost and, therefore, search "actively". For high-skilled workers, instead, on top of the active search, there is also a possibility of "passive" search. A worker is searching passively if she agrees to consider an offer when a headhunter calls. Screening and passive search match exactly the two main features of the headhunter industry. To study the third feature of headhunters, the role in wage setting, I use period-by-period wage bargaining for wage determination. I then isolate the effects of matching on wage distribution by closing the wage bargaining channel.

Having set up the model, I apply the following calibration strategy. First, I calibrate the model without headhunters to match moments of the wage distribution and aggregate labor market moments in the U.S. in the 1970s. The key calibrated parameters include those characterizing the exogenous distributions of workers over skills and firms over productivity. The U.S. labor market in the 1970s is well approximated by the model with no, or limited, role for headhunters. Having fixed the parameters not related to the headhunters, I then introduce the headhunter channel to the model and calibrate the related parameters to target the moments of the headhunter industry in the 2010s as well as the change in the joint distribution of top workers and firms between the 1980s and 2000s[6]. At the same time, I introduce skill-biased technological change to match the increase in the 90/50 wage ratio from 1970 to 2010. I do it by increasing the degree of complementarity in the production function[7]. Having both mechanisms in the model

[5] Other studies including two-sided heterogeneity in the labor market include Shimer and Smith (2000); Postel-Vinay and Robin (2002); Teulings and Gautier (2004); Gautier and Teulings (2015) and many others.

[6] The main quantitative results of the paper are not a direct consequence of targeting the change in the joint distribution in the calibration of the headhunter channel. The robustness analysis presented in the Appendix shows that different calibration strategies that do not rely on the change in the joint distribution lead to a greater role of the headhunter channel in explaining the rise of top wages. Matching the observed change in the joint distribution lowers the skill threshold for workers being eligible for the headhunter channel and reduces the average quality of matches at the top, reducing the top wages and, therefore, top wage shares.

[7] In the functional form for the production function chosen in the calibration, the parameter responsible for the degree of complementarity also changes the relative productivity of high-skilled workers. When the degree of complementarity increases, so does the productivity of high-skilled workers relative to the low-skilled workers, hence it is able to replicate the changes in the productivity attributed to the skill-biased technological change. When I increase the degree of complementarity, I also increase the search

allows me to evaluate the relative contribution of each mechanism to the change occurring on different parts of the wage distribution.

My calibration strategy answers the question how would the distribution of wages (and, therefore, also the top wages) have changed between the 1970s and 2010s if skill-biased technological change and headhunters had been the only factors raising top wage inequality. To assess the relative contribution of the two factors, I then shut down one channel at a time: the skill-biased technological change or the headhunter channel. I also give a chance to skill-biased technological change to explain all the increase in wage inequality without the headhunters. To do that, I change the increase in the degree of complementarity to match the increase in the 90/50 wage ratio and assess how the model fits other moments. To further study the mechanisms I then exploit the richness of the model and compare several statistics related to the wage distribution with and without the headhunters. Importantly, I perform experiments in line with Song et al. (2019) and compare results from the model-generated data to the U.S. data. This allows me to see whether the improvement in matching in the model has similar features to the observed improvement.

The main quantitative result of the paper is that the rise of headhunters accounts for 32% of the increase in the top 10% wage shares in the U.S. from the 1970s to 2010s. Out of 32%, the improvement in the matching between workers and firms accounts for 19% and the change in wages due to the improvements in outside options accounts for the other 13%. Skill-biased technological change contributes to another 23% of the top 10% wage share increase, and the interaction between the two factors raises the top 10% share by 3%[8]. The sharp increase of top wages in the model is mainly due to improved matching after the introduction of headhunters. Comparing joint distributions of worker-firm matches in the two steady states reveals a pattern similar to empirical results of Song et al. (2019)[9], where most types of firms lose the highest-skilled workers and where the highest-paying firms gain those workers disproportionately. The headhunter channel generates the strong non-linearity in the change in assortative matching observed in the data, with a disproportionate improvement in matching for highest-skilled workers. I am not aware of other theoretical models able to generate such non-linearity. If I allow the model to match the increase in the 90/50 wage ratio without the headhunter channel, the model explains a smaller share of the increase in the top 10% wage share but

costs proportionally, so that the sorting is not affected much by this change. If the search costs stayed the same, the quit rate would have increased to unrealistic levels while there was no positive trend in the observed quit rate over the period studied in the paper.

[8]The numbers for the top 1% wage share are 10% (3.5% and 6.5%), 10% and 1%, respectively.

[9]Estimating a fixed effect regression is only one way to evaluate sorting in the labor market. Eeckhout and Kircher (2011) show potential problems with identifying sorting with estimated fixed effects. Studies using non-parametric techniques, instead, find a higher degree of sorting than in the studies using fixed effects regressions. Hagedorn, Law and Manovskii (2017) find the correlation between worker and firm ranks to be 0.75 in Germany, Lise, Meghir and Robin (2016) who find significant sorting for college-graduates in the U.S., and Borovičková and Shimer (2017) who find a correlation between 0.4 and 0.6 in Austria. Schulz and Lochner (2016) show, using non-parametric techniques, that sorting increased in Germany between 1998 and 2008.

overshoots the 50/10 wage ratio. This happens exactly because of the absence of a strong non-linearity of the skill-biased technological change. An increase in complementarity shifts the whole distribution of wages to the right. The non-linearity generated by the headhunters allows shifting the right tail of the distribution farther apart from the rest of the distribution without changing the shape of the distribution in the middle[10].

The model relates to other theoretical models showing the importance of assortative matching for wage distribution. Bagger and Lentz (2019) is the closest study. They show that on-the-job search is a crucial mechanism to generate assortative matching in a Diamond-Mortensen-Pissarides model with two-sided heterogeneity. Bagger and Lentz (2019) consider only active on-the-job search. Uren and Virag (2011), instead, show that skill requirements are important to generate an increase in between-group inequality (increased differences between wages of workers with different skill level). Skill requirements play a similar role as does the screening by headhunters. Uren and Virag (2011) study the overall shape of the wage distribution, while this paper focuses on headhunters and the top part of the distribution.

This paper presents two blocks of independent empirical evidence supporting the mechanism. First, it uses micro-level data on CEO compensation of listed companies in the U.S. to study the effects of a change of the CEO on CEO compensation in the company. The main result is that firms pay significantly more to new CEOs comparing to the previous ones, and this difference is higher during the periods when headhunters are used more intensively and in the states where there are less legal obstacles to the activity of headhunters[11]. These results fit the prediction of the model that headhunters improve matching between firms and CEOs and, therefore, increase wages at the top. Second, the paper uses cross-country differences in the use of headhunters in Europe in 1997 to show that in countries where headhunters were used to a larger extent the top income shares increased by more in the following years. This evidence is in line with the prediction of the model that the more the high-productive firms use headhunters, the better is the improvement of matching at the top, and the higher are the top wages.

The remainder of the paper is organized as follows. Section 2 presents the theoretical model. Section 3 presents the quantitative results. Section 4 discusses the available empirical evidence about headhunters and headhunter industry. Section 5 presents suggestive empirical evidence. Section 6 concludes.

[10]The rise of headhunters can be also viewed as a reason for the shift in the mean income growth rates for high-skilled workers in the model of Gabaix et al. (2016). Gabaix et al. (2016) introduce an exogenous increase in mean income growth rate for some workers in 1980s, motivated by globalization and technological change. Headhunters, who allow high-skilled workers to work for the top firms, generate the increase in income growth rate for high-skilled workers due to better matches.

[11]The legal obstacles are proxied by the enforceability of non-compete agreements as proposed by Garmaise (2011).

2 Model

2.1 Environment

The economy is populated by a continuum of workers and firms. Workers differ in their skill level, e, and supply one unit of labor if employed. When a worker is unemployed she receives an unemployment benefit, $b(e)$. Firms differ in their productivity level, p. Each firm can hire one worker. Both workers and firms discount their future utility with a discounting rate β.

There are two channels for matching workers and firms. First, there is the standard, or vacancy, channel where every worker and firm can participate by paying a per-period cost. Second, there is the headhunter channel where every firm and only high-skilled workers, those with skill above an exogenous threshold, $\hat{e}$, can participate. To participate in the headhunter channel firms have to pay a per-period cost, while workers pay the search cost only if they are matched. Workers and firms participating in each channel are randomly matched by a standard CRS matching technology.

All workers, unemployed and employed, can search for a job. Each period workers decide whether to search for a job through vacancies (search actively) and/or to be available for a headhunter company (search passively) if her skill is higher than a threshold $\hat{e}$. Firms, instead, can choose only one of the channels to search for a worker. The production of a match depends on the firm's productivity level and the worker's skill level via a supermodular production function. The wage in a match is determined period by period via bargaining. Separation of matches depends on two factors: idiosyncratic exogenous separation shock, s; and endogenous worker's quit rate, $s_Q(.)$. There is no aggregate uncertainty.

2.2 Timing

The time is discrete. Inside every period, first, exogenous separations happen. Then, workers and firms decide in which markets to participate, and new firms decide whether to enter the market. After that, workers searching for a job and firms searching for a worker are matched. Finally, existing matches produce and wages and unemployment benefits are paid.

2.3 Matching

The two channels are labeled as the vacancy, V, and the headhunter, H, channels. In channel $i \in \{V, H\}$ workers and firms meet via a standard CRS matching technology: $m_i = m_i(u_i + a_i, v_i)$, where m_i is the number of matches, u_i and a_i are the numbers of unemployed and employed workers participating in the channel, respectively, and v_i is the number of firms participating in the channel. Define the market tightness of channel i, θ_i, as $\theta_i = \frac{v_u}{u_i+a_i}$. The job finding rate for a worker using channel i is $f_i(\theta_i) = \frac{m_i(u_i+a_i,v_i)}{u_i+a_i} = m_i(1, \theta_i)$ and the job filling rate for a firm is $q_i(\theta_i) = \frac{m_i(u_i+a_i,v_i)}{v_i} = \frac{m_i(1,\theta_i)}{\theta_i}$.

2.4 Production Technology

The production of a match between a worker with skill e and a firm with productivity p is determined by a function $y(e,p)$. This function is increasing and quasi-concave in both components, that is $y'_p > 0$, $y'_e > 0$, $y''_{pp} \leq 0$ and $y''_{ee} \leq 0$. Moreover, $y(e,p)$ has a property of supermodularity having positive cross-derivatives: $y''_{ep} > 0$, $y''_{pe} > 0$. Supermodularity is necessary for complementarity between the worker's skill and the firm's productivity that creates incentives for positive assortative matching.

2.5 Worker Problem

Every worker can be either employed or unemployed. An unemployed worker with skill e consumes the unemployment benefit, $b(e)$, and searches for a job in the next period. The value of unemployment, $U(e)$, can be written as:

$$U(e) = b(e) + \beta\left(U(e) + S_U(e)\right), \tag{1}$$

where $S_U(e)$ is the value of search for an *unemployed* worker.

A worker with skill e employed in a firm with productivity p consumes the wage, $w(e,p)$, this period, and next period the match can be exogenously separated with probability s, in which case the worker becomes unemployed, or with probability $(1-s)$ the match survives and the worker can continue to search on-the-job. The value of work, $W(e,p)$, is:

$$W(e,p) = w(e,p) + \beta\left(sU(e) + (1-s)\left(W(e,p) + S_E(e,p)\right)\right), \tag{2}$$

where $S_E(e,p)$ is the value of search for an *employed* worker.

The value of search is different for workers with different skill level as only the high-skilled workers have a chance to be contacted by a headhunter. Consider first the problem of a low-skilled unemployed worker. The *low-skilled unemployed worker* is excluded from the headhunter channel so the only choice that she has is between searching through the vacancy channel and not searching. Low-skilled worker's value of search can be written as:

$$S_U(e) = \max\left\{S_{UV}(e), 0\right\}, \ \ if \ \ e < \hat{e}, \tag{3}$$

where $S_{UV}(e)$ is the value of search through the vacancy channel for an unemployed worker.

For the *high-skilled unemployed worker*, the problem is the same but she chooses among four options: search through vacancies, wait for a headhunter call, do both, or be inactive. The value of the search of a high-skilled unemployed worker can be written as:

$$S_U(e) = \max\left\{S_{UV}(e), S_{UH}(e), S_{UVH}(e), 0\right\}, \ \ if \ \ e \geq \hat{e}, \tag{4}$$

where $S_{UH}(e)$ is the value of search through the headhunter channel and $S_{UVH}(e)$ is the value of search through both channels for an unemployed worker.

When an unemployed worker is searching through the vacancy channel, with probability $f_V(\theta_V)$ she will receive an offer from a firm with productivity p that will be drawn from a distribution of firms posting vacancies. The worker then decides whether to accept the offer and receive the difference between the value of employment in this firm, $W(e,p)$, and unemployment, $U(e)$, or to stay unemployed and have no gain. To participate in the channel, the worker has to pay the search cost, $c_{wV}(e)$, every period of active search. Therefore, the value of search through the vacancy channel for an unemployed worker is the following:

$$S_{UV}(e) \equiv f_V(\theta_V) E_{p|V}[\max\{W(e,p), U(e)\} - U(e)] - c_{wV}(e).$$

The value of search through the headhunter channel, or passive search, differs in four respects: offer arrival probability, $f_H(\theta_H)$; search cost, $c_{wH}(e)$; the search cost is paid only if the offer arrives; and the offer is drawn from a different distribution (distribution of firms using headhunters). It is assumed that every eligible worker decides whether to agree to consider an offer in case of a headhunter's call in the beginning of the period before the offer has materialized. The worker will have to pay the search cost (spend time on the interviews or risk being penalized by current employer) only if she receives the call that period. The value of search through the headhunter channel is the following:

$$S_{UH}(e) \equiv f_H(\theta_H)\left(E_{p|H}[\max\{W(e,p), U(e)\} - U(e)] - c_{wH}(e)\right).$$

The value of search through both channels is just a combination of the two, with an implicit assumption that better firms are using the headhunter channel[12]. The value function is the following:

$$\begin{aligned} S_{UVH}(e) \equiv \; & f_H(\theta_H)\left(E_{p|H}[\max\{W(e,p), U(e)\} - U(e)] - c_{wH}(e)\right) \\ & + f_V(\theta_V)(1 - f_H(\theta_H)) E_{p|V}[\max\{W(e,p), U(e)\} - U(e)] \\ & - c_{wV}(e). \end{aligned}$$

Consider an *employed worker*. She also decides whether to participate in the channels but has a different outside option. Because the worker can always stay in the current firm, the value of search now depends also on the productivity of the current employer, p. Similarly to a low-skilled unemployed worker, a *low-skilled employed worker* may choose between searching through the vacancy channel and not searching at all, with the value of search being:

$$S_E(e,p) = \max\{S_{EV}(e,p), 0\}, \; if \; e < \hat{e}, \tag{5}$$

where $S_{EV}(e,p)$ is the value of search through the vacancy channel for an employed worker.

A *high-skilled employed worker* may choose again among four options: search through the vacancy channel, search through the headhunter channel, search through both channels, or not search at all. The value of search can be written as:

$$S_E(e,p) = \max\{S_{EV}(e,p), S_{EH}(e,p), S_{EVH}(e,p), 0\}, \; if \; e \geq \hat{e}, \tag{6}$$

[12]This assumption will be satisfied in the equilibrium.

where $S_{EH}(e,p)$ is the value of search through the headhunter channel and $S_{EVH}(e,p)$ is the value of search through both channels for an employed worker.

The values of search through a particular channel differ from the ones for an unemployed worker due to a different outside option - an employed worker can always stay with the current employer if the new match is with a less productive firm. The three values of search, therefore, are:

$$\begin{aligned}
S_{EV}(e,p) \equiv\ & f_V(\theta_V) E_{p'|V}\left[\max\{W(e,p'),W(e,p)\} - W(e,p)\right] - c_{wV}(e),\\
S_{EH}(e,p) \equiv\ & f_H(\theta_H)\left(E_{p'|H}\left[\max\{W(e,p'),W(e,p)\} - W(e,p)\right] - c_{wH}(e)\right),\\
S_{EVH}(e,p) \equiv\ & f_H(\theta_H)\left(E_{p'|H}\left[\max\{W(e,p'),W(e,p)\} - W(e,p)\right] - c_{wH}(e)\right)\\
& + f_V(\theta_V)(1 - f_H(\theta_H)) E_{p'|V}\left[\max\{W(e,p'),W(e,p)\} - W(e,p)\right]\\
& - c_{wV}(e).
\end{aligned}$$

2.6 Firm Problem

Firms with vacant positions also need to choose a channel through which to find a worker. Unlike workers, all firms solve the same problem (regardless of their productivity level) and they can choose only one channel. The value of a vacant job is defined as:

$$V(p) = \max\{V_V(p); V_H(p)\}, \tag{7}$$

where $V_V(p)$ is the value of hiring through the vacancy channel and $V_H(p)$ is the value of hiring through the headhunter channel.

If the firm decides to post a vacancy, it pays the per-period cost $c_{fV}(p)$ and is matched with a worker with probability $q_V(\theta_V)$. The worker will be drawn from the distribution of workers searching through the vacancy channel. The worker will accept the new match with probability $P(A)$, where A is the event of a worker with skill e who is unemployed or employed in a firm with productivity p' accepting the match with the new firm. This happens when the worker doesn't have a better offer in the same period and if she works in a firm with lower productivity[13] (if searching on-the-job). The probability is either 0 or 1 for each worker in the pool of potential matches. If the match is formed, the firm receives the difference between the value of a job with a worker with skill e, $J(p,e)$, and the value of a vacancy. The value of hiring through the vacancy channel for a firm is:

$$V_V(p) = -c_{fV}(p) + \beta\left(V(p) + q_V(\theta_V) E_{e|V}\left[P(A)(J(p,e) - V(p))\right]\right). \tag{8}$$

Similarly, if a firm decides to hire through the headhunter channel, it pays the per-period cost $c_{fH}(p)$ and is matched with a worker with probability $q_H(\theta_H)$. The worker will be drawn from the distribution of workers searching through the headhunter channel. The value of hiring through the headhunter channel is:

$$V_H(p) = -c_{fH}(p) + \beta\left(V(p) + q_H(\theta_H) E_{e|H}\left[P(A)(J(p,e) - V(p))\right]\right). \tag{9}$$

[13] As each firm hires only one worker, the internal promotion cannot be modeled explicitly. One could interpret the hires through the vacancy channel as internal promotions, in this case, the worker accepts the new position if the position is of a higher rank than her current position. Even though the firms in the model have different identities they can be a part of a large corporation that owns these firms/positions.

The value of a job is standard. The firm receives the product of the match, pays the wage and in the next period the match may be separated due to an exogenous shock or due to an endogenous worker's quit to another firm. The value of a job can be written as:

$$\begin{aligned} J\left(p,e\right) = \; & y\left(e,p\right) - w\left(e,p\right) + \\ & +\beta\left(\left(s + s_Q\left(.\right)\left(1-s\right)\right)V\left(p\right) + \left(1 - s_Q\left(.\right)\right)\left(1-s\right)J\left(p,e\right)\right). \end{aligned} \tag{10}$$

There is an ex-ante free entry condition. Firms do not know their level of productivity before entering the market. The firm draws its productivity from an exogenous distribution after paying the entry cost, F.The free entry condition for the firms is the following:

$$E_p\left[V\left(p\right)\right] = F. \tag{11}$$

2.7 Wages

Wages are determined via period-by-period wage bargaining between the workers and the firms. Wage in a match between a worker with skill e and a firm with productivity p is the solution to the bargaining problem[14]:

$$w\left(e,p\right) = \max_w \left(\hat{W}\left(e,p,w\right) - U\left(e\right)\right)^{\vartheta}\left(\hat{J}\left(e,p,w\right) - V\left(p\right)\right)^{1-\vartheta},$$

where ϑ is the bargaining power of the worker, and functions $\hat{W}$ and $\hat{J}$ are such that $\hat{W}\left(e,p,w\left(e,p\right)\right) = W\left(e,p\right)$ and $\hat{J}\left(e,p,w\left(e,p\right)\right) = J\left(e,p\right)$. The solution to the problem must satisfy the standard sharing rule for each match:

$$\vartheta\left(\hat{J}\left(e,p,w\right) - V\left(p\right)\right) = \left(1-\vartheta\right)\left(\hat{W}\left(e,p,w\right) - U\left(e\right)\right).$$

2.8 Steady-State Separating Equilibrium

In this section I discuss a particular structure of the equilibrium. Given supermodularity of the production function, the most reasonable equilibrium is the one where high-productive firms would hire through the headhunter channel, while low-productive firms would use the vacancy channel. Such equilibrium requires some assumptions on the cost functions, production function, and initial distributions of workers and firms. I discuss these assumptions in the Appendix. It is assumed throughout the section that such assumptions hold. In the numerical exercises, I verify that such assumptions do hold under the baseline calibration.

[14]In the simulations, I add a failed bargaining cost to the firms so that all the matches are formed. In the equilibrium, less than 0.1% of matches are enforced in this way and these matches are with the lowest-skilled workers. It has a negligible effect on the results and is done solely for computational purposes.

2.8.1 Distributions

First, I need to specify distributions that will be used in expectations. Let $F(p)$ be the initial distribution of firm productivity and $G(p)$ the measure of firms with an open vacancy, both have support $\left[\underline{p}, \overline{p}\right]$. Denote as $\hat{p}$ the cutoff level of firm productivity such that firms with productivity above $\hat{p}$ hire through the headhunter channel and firms below $\hat{p}$ hire through the vacancy channel. Also, let $H(e)$ be the initial distribution of workers over skill, $L_V(e)$ be the measure of employed workers searching for a job through the vacancy channel, $L_H(e)$ the measure of employed workers searching for a job through the headhunter channel, $L_{VH}(e)$ the measure of employed workers searching for a job through both channels, and $U(e)$ the measure of unemployed workers over the skill level (all with support $[\underline{e}, \overline{e}]$). Finally, let $\Phi(e,p)$ be the joint measure of active matches and $\Lambda_i(e,p)$ be the measure of active matches in which a worker is searching for a new job through channel $i \in \{V, H, VH\}$.

2.8.2 Workers

Given the structure of the equilibrium under consideration and the distributions defined above, we can now specify the expectations.

Under our assumptions, low-skilled workers are excluded from the headhunter channel so they can search only through the vacancy channel. The value of search is:

$$S_U(e) = S_{UV}(e) \equiv f_V(\theta_V) \int_{\underline{p}}^{\hat{p}} (W(e,p) - U(e))\, dG(p) - c_{wV}(e). \tag{12}$$

For high-skilled unemployed workers it is optimal to search through both channels. Their value of search is then:

$$S_U(e) = S_{UVH}(e), \tag{13}$$

and the exact expression for $S_{UVH}(e)$ is defined in the Appendix. Under our assumption, better firms use the headhunter channel. If an unemployed high-skilled worker receives an offer through the headhunter channel she will accept it regardless of receiving an offer through the vacancy channel or not. Instead, this worker will accept an offer from the vacancy channel only if she doesn't receive an offer through the headhunter channel.

For employed workers the value of search is the same as the value of search for unemployed, except from the outside option. The value function of search will be, as before:

$$S_E(e,p) = \max\{S_{EV}(e,p); 0\}.$$

We can define the value of search through the vacancy channel for a low-skilled employed worker as:

$$S_{EV}(e,p) \equiv f_V(\theta_V) \int_{\underline{p}}^{\hat{p}} \max\{W(e,p') - W(e,p); 0\}\, dG(p') - c_{wV}(e). \tag{14}$$

Because the worker will accept offers only from more productive firms, the value can be rewritten as:

$$S_{EV}(e,p) \equiv f_V(\theta_V) \int_p^{\hat{p}} (W(e,p') - W(e,p))\, dG(p') - c_{wV}(e).$$

To search from employment, the value of search for a worker with skill e and working in a firm with productivity p must be positive:

$$S_{EV}(e,p) \geq 0.$$

This equation (when satisfied with equality) implicitly determines the level of the firm productivity such that a worker with a skill level e does not search for a new job: $\tilde{p}_V(e)$ (for $e < \hat{e}$). If a worker with skill e works in a firm with productivity below $\tilde{p}_V(e)$, she searches for another job and doesn't search otherwise.

For a high-skilled employed worker the value of search consists of four options but in this structure of equilibrium one of them (searching only through vacancies) will never be optimal[15]. The value of search can be defined as:

$$\begin{aligned} S_E(e,p) &= \max\{S_{EV}(e,p); S_{EH}(e,p); S_{EVH}(e,p); 0\} \\ &= \max\{S_{EH}(e,p); S_{EVH}(e,p); 0\}. \end{aligned}$$

For a high-skilled worker with skill level e there are now two cutoff productivity levels $\tilde{p}_{VH}(e)$ and $\tilde{p}_H(e)$, with $\tilde{p}_H(e) \geq \tilde{p}_{VH}(e)$. If the worker is employed in a firm with productivity below $\tilde{p}_{VH}(e)$ she will search for another job through both channels. If she works in a firm with productivity level between $\tilde{p}_{VH}(e)$ and $\tilde{p}_H(e)$, she will search only through the headhunter channel. If she works in a firm with productivity above $\tilde{p}_H(e)$, she will not search for another job at all. Before defining the conditions that determine these cutoffs, we need to define the value functions.

The value of search through the headhunter channel for a high-skilled worker can be defined as:

$$S_{EH}(e,p) \equiv f_H(\theta_H) \left(\int_{\max\{\hat{p};p\}}^{\overline{p}} (W(e,p') - W(e,p))\, dG(p') - c_{wH}(e) \right). \tag{15}$$

The value of search through both channels is defined in the Appendix.

It is easy to show that, given e, $S_{EVH}(e,p)$ is higher than $S_{EH}(e,p)$ for small p, but $S_{EVH}(e,p)$ decreases faster, so they will always have just one intercept. The equality:

$$S_{EVH}(e,p) = S_{EH}(e,p)$$

defines the cutoff productivity level of searching through both channels for each worker type, $\tilde{p}_{VH}(e)$, while the equality

$$S_{EH}(e,p) = 0$$

defines the cutoff productivity level for searching only through the headhunter channel, $\tilde{p}_H(e)$.

The value functions of work and unemployment are defined as before.

[15] See Appendix.

2.8.3 Firms

We can also rewrite the values of hiring through the vacancy and the headhunter channels given distributions defined above. The exact expressions are presented in the Appendix.

One can show that under reasonable conditions on the values of the search costs and production function, the value of hiring through a headhunter, $V_H(p)$, is lower than the value of posting a vacancy, $V_V(p)$, for small p but $V_H(p)$ is increasing faster with p[16]. Thus, there will be only one intercept between $V_H(p)$ and $V_V(p)$, an endogenous threshold $\hat{p}$, such that

$$\max\{V_V(p); V_H(p)\} = V_V(p)$$

for $p < \hat{p}$ and

$$\max\{V_V(p); V_H(p)\} = V_H(p)$$

for $p > \hat{p}$. The cutoff productivity is determined by

$$V_V(\hat{p}) = V_H(\hat{p}).$$

Finally, the value of an active match for a firm is defined as before, with the exact expression for the quit rate defined in the Appendix.

2.8.4 Equilibrium

The steady-state equilibrium, given the initial distributions of workers over skill and firms over productivity, the exogenous skill threshold, the matching functions, and the production function, is defined by the value functions, the endogenous distributions, the decision rules, and the wage function. The decision rules must be consistent with the value functions. The value functions must be consistent with the endogenous distributions and the wage function. Endogenous distributions must satisfy the balances given the decision rules. The wage function must solve the sharing rule given the value functions.

The balances guarantee that the equilibrium distribution is stationary over time. In the equilibrium, the inflow of workers to every worker-firm distribution bin must be equal to the outflow of workers from that bin. The equilibrium density of active matches for a pair of workers with skill e and firms with productivity p, $\phi(e,p)$, must satisfy:

$$\phi(e,p)(s + s_Q(e,p)(1-s)) = i_E(e,p) + i_U(e,p), \tag{16}$$

where the left-hand side is the total outflow from active matches (exogenous plus endogenous separations), and the right-hand side is the total inflow into the matches from employment, $i_E(e,p)$, and unemployment, $i_U(e,p)$. The inflow rates are defined in the Appendix.

[16]See Appendix.

2.9 Solution Method

To find a steady state equilibrium I first guess the decision rules. Given the decision rules, I solve the system of balances equations using non-linear solution methods (trust-region or Broyden methods). The solution to the system is the stationary distribution of active matches. Then I compute the rest of endogenous distributions given the distribution of active matches and exogenous initial distributions of worker and firm types. Given distributions, I solve for the value functions for workers and firms and the wage function using non-linear solution methods. Finally, I compute new decision rules based on the value functions. I iterate these steps until convergence.

2.10 Idiosyncratic Headhunter Costs

The most important extension of the model is introduced to capture the fact that not all firms hire employees for top positions through headhunters in the data. In the baseline model, every firm above the threshold $\hat{p}$ hires through the headhunters. I introduce an additional idiosyncratic cost for hiring through the headhunter channel. Every firm with productivity above $\hat{p}$ and an open position draws a cost c_{fN} that is paid only if the firm wants to hire through the headhunter channel[17]. This cost might reflect corporate practice, an existence of a preferred candidate inside the firm, or specificity of the position. A firm with a high cost will have to post a vacancy even if it would prefer to hire through the headhunter channel absent the cost. Updated value of an open position, $\tilde{V}(p)$, can be written as:

$$\tilde{V}(p) = \max\left\{V_V(p); V_H(p) - c_{fN}\right\}.$$

This extension doesn't alter the model significantly[18] but brings it closer to the data. Because the proportion of top firms using headhunters depending on productivity is unobservable, the distribution of the costs will be chosen to target a probability of a top firm (that would hire through the headhunter channel without the idiosyncratic cost) hiring through the headhunter channel. To allow for some flexibility, I split the firms into two groups with different probability of hiring through the headhunter channel. The firms in the top 10% of productivity distribution hire through the headhunter channel with probability χ_{10} and the firms in 10-20% of productivity distribution with probability χ_{20}.

2.11 Calibration

The model is calibrated to monthly frequency. The calibration strategy is the following. First, the version of the model without the headhunter channel is calibrated to match the labor market in the U.S. in the 1970s. Then the parameters related to the headhunter channel are calibrated to match the moments of the headhunter industry in the U.S. in

[17]It is assumed that this cost is always large enough for firms with productivity below $\hat{p}$ so that the firm will never choose to hire through the headhunter channel.

[18]Equations for the value functions and the balances are presented in the appendix.

the 2010s and the change in the joint distribution of top workers' and firms' fixed effects observed in the data. To take into account the skill-biased technological change from the 1970s to 2010s, I also change the degree of complementarity in the production function to match the increase in the 90/50 wage ratio between the 1970s and 2010s.

To calibrate the model, I need to specify the exogenous distributions of workers, $H(e)$, and firms, $F(p)$, the productivity function, $y(e,p)$, the matching function, $m(u,v)$, the skill threshold, $\hat{e}$, the search costs, $c_{fH}(p)$, $c_{fV}(p)$, $c_{wH}(e)$, $c_{wV}(e)$, and the distribution of the idiosyncratic cost c_{fN}. The functional form for the initial distributions of firms and workers is chosen to be beta with the parameters λ_{F1} and λ_{F2} for the firms, and λ_{W1} and λ_{W2} for the workers. The beta distribution is over the support $[0,1]$ that is not suitable for the experiment. I rescale the support of the distribution to be $[1,\overline{p}]$ and $[1,\overline{e}]$ with $\overline{p}=\overline{e}$.

I choose the following functional forms. The matching function has the standard Cobb-Douglas form:

$$m_i(u,v) = M_i u^{\sigma} v^{1-\sigma} \text{ for } i \in \{V,\, H\},$$

with a standard value for the matching elasticity of 0.5. The production function has the form[19]:

$$y(e,p) = (e \cdot p)^{\gamma},$$

with normalization $\gamma = 1$ in the 1970s. The cost functions have the following forms:

$$\begin{aligned} c_{fH}(p) &= c_{fH} \cdot p^{c_f} \\ c_{fV}(p) &= c_{fV} \cdot p^{c_f} \\ c_{wH}(e) &= c_{wH} \cdot e^{c_w} \\ c_{wV}(e) &= c_{wV} \cdot e^{c_w} \end{aligned}$$

with $c_f = 1.85$ and $c_w = 0.8$ chosen to fit reasonable on-the-job search strategies. Finally, unemployment benefits are:

$$b(e) = b \cdot e^{b_w}$$

with $b_w = 0.8$.

I use standard values for the discount factor (0.99) and the bargaining power (0.6). All other parameters not related to the headhunter channel are calibrated to match the wage distribution and other moments of the labor market in the 1970s. There are nine parameters to calibrate for the steady state without the headhunters: the parameters of the distribution, λ_{F1}, λ_{F2}, λ_{W1} and λ_{W2}, the maximum type, $\overline{p}$ or $\overline{e}$, the exogenous separation rate, s, the matching function efficiency in the standard channel, M_V, and the vacancy channel search costs for workers, c_{wV}, and for firms, c_{fV}. Eight targets are chosen to jointly set these nine parameters: the top 1% and the top 10% wage shares, the 90/50 wage ratio, the 50/10 wage ratio, the unemployment rate, the job finding rate, the quit rate, and the estimate of vacancy cost relative to annual worker's wage. The parameters are jointly calibrated to match the targets.

[19] It is easy to see that this production function is supermodular with $\frac{\partial^2 y(e,p)}{\partial e \partial p} = \frac{\partial^2 y(e,p)}{\partial p \partial e} = \gamma^2 (e \cdot p)^{\gamma-1} > 0$.

The targets for the calibration are taken from the following sources. The top 1% and the top 10% wage shares are from Piketty (2014). The 90/50 and the 90/10 wage ratios are computed using the data from Song et al. (2019). The unemployment rate is from BLS. The quit rate is taken from NBER Macrohistory Database, the data is for the 1960s but is taken as a proxy for the 1970s. The job finding rate is from Shimer (2005). The vacancy costs estimates are from Manning (2011).

There are six parameters related to the headhunter channel to which I need to assign values: the search cost for workers, c_{wH}; the search cost for firms, c_{fH}; the skill threshold, $\hat{e}$; the share of firms using headhunters in top 10%, χ_{10}, and 10-20%, χ_{20}[20], and the relative matching efficiency of the headhunter channel, M_H/M_V. I calibrate these parameters to match six targets: the estimate of the positive response rate by managers to a call by a headhunter; the average fee of headhunters; and the change in the joint distribution of top workers and firms. The targets for the response rate and the average fee are taken from Cappelli and Hamori (2013)[21]. Cappelli and Hamori (2013) estimate that around 50% of managers say "yes" when a headhunter calls and asks if the manager is willing to consider an offer. I use the positive response rate to calibrate the workers' passive search cost. This search cost determines the workers' search strategies in the model. I calibrate the cost so that 50% of high-skilled workers (with the skill above the screening threshold) choose to search passively. Cappelli and Hamori (2013) also present the evidence for the size of the average headhunter fee. They show that the average fee equals to around 30% of the annual wage of the hired worker. I calibrate the firms' headhunter cost to match this number on average. The targets for the change in the joint distribution of top workers and firms are from Song et al. (2019) who estimate a change in the joint distribution of workers' and firms' fixed effects between 1980s and 2000s. I use the change in the joint distribution of top two bins of workers and firms. This gives me four more targets needed: change in the share of top 10% workers in top 10% firms (F10-W10); change in the share of 10-20% workers in top 10% firms (F10-W9); change in the share of 10-20% workers in 10-20% firms (F9-W9); and change in the share of top 10% workers in 10-20% firms. As the parameters of the headhunter channel are not precisely estimated in the data, I present results for alternative calibrations of the headhunter channel in the Appendix. On top of the headhunter channel, I also increase the degree of complementarity, γ, in order to match the change in the 90/50 wage ratio between the 1970s and the 2010s.

The results of the calibration are presented in Table 1. The model matches well the main characteristics of the wage distribution and the labor market in the 1970s. The targets of the headhunter channel are also matched reasonably well. Of particular interest, the screening threshold is 2.7. The headhunters target only workers who are at least 2.7 times more productive than the least skilled workers. 16.9% of workers fall above this threshold. The workers' search cost through the headhunter channel is quite high, the average cost of participating employed workers corresponds to 80% of their annual

[20] It is equivalent to calibrating type-specific distribution functions for the idiosyncratic cost c_{fN}.

[21] Cappelli and Hamori (2013) use the data for executives working in the financial services industries in New York area. It is hard to say whether the data is representative of the other industries covered by the headhunters but it is the only available source of data of such quality.

wage (but less than 10% of their value of work). But because the cost is paid only when the worker is contacted, the expected cost of participating in the channel is only around 1.4% of the annual wage. The high cost also highlights the tremendous gains for the workers hired through the headhunter channel that are even higher than the cost.

Table 1: Calibrated Parameters

Parameter	Value	Target	Data	Model
Wage distribution, 1970s				
Beta parameter, λ_{F1}	1.55	Top 1% wage share	5.1%	4.41%
Beta parameter, λ_{F2}	18	Top 10% wage share	25.7%	25.86%
Beta parameter, λ_{W1}	2.15	90/50 wage ratio	2.06	2.21
Beta parameter, λ_{W2}	33	50/10 wage ratio	2.03	1.96
Maximum types, $\overline{p}$, $\overline{e}$	25	-	-	-
Labor market, 1970s				
Separation rate, s	0.027	Unemployment rate	5%	5%
Matching function, M_V	0.35	Job finding rate	50%	50%
Search cost - vacancies, c_{wV}	2.4	Quit rate	2%	1.98%
Vacancy cost, c_{fV}	0.15	Vacancy cost estimates	8%	7.99%
Headhunter industry, 2010s				
Headhunter search cost, c_{wH}	17	Positive response rate	50%	56.7%
Headhunter firm cost, c_{fH}	0.63	Headhunter average fee	30%	30.8%
Screening threshold, $\hat{e}$	2.7	ΔF10-W10	0.0129	0.0137
Share of top 10% firms using HH, χ_{10}	0.3	ΔF10-W9	0.0085	0.0087
Share of 10-20% firms using HH, χ_{20}	0.15	ΔF9-W9	0.0031	0.0042
Reltive matching efficiency of HH, M_H/M_V	0.21	ΔF9-W10	0.0059	0.0059
Skill-biased technological change, 1970s-2010s				
Degree of complementarity, γ	1.09	Δ 90/50 ratio	0.52	0.53

3 Results

3.1 Inequality

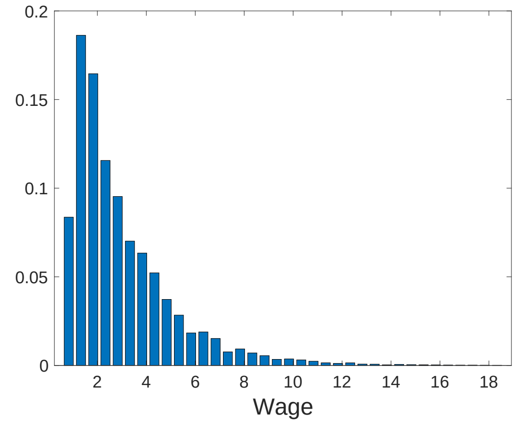
The headhunter channel changes the wage distribution. Without the headhunter channel, the distribution has a peak close to the minimal possible wage and then decreases, having a form close to Pareto (Figure 1a). When the headhunter channel is present in the model, the distribution still has a similar form, but it has a fatter right tail (Figure 1b). The headhunter channel generates the fat tail of the wage distribution in this model. The reason for this is the following. Without the headhunter channel the probability of matching a high-skilled worker with a high-productive firm is lower than matching a high-skilled worker with a low-productive firm (due to the fact that there are relatively few high-productive firms), so there will be large shares of high-skilled workers working in low-productive firms and low-skilled workers in high-productive firms. Wages of low-skilled workers are lower than wages of high-skilled workers in the same type of firm. And because only some high-productive firms will be matched with high-skilled workers without the headhunter channel, there will be a small mass of workers getting very high wages. When, instead, there is a possibility to hire only high-skilled workers through the headhunter channel, high-productive firms will be matched only with high-skilled workers and all of them will receive relatively high wages; this corresponds to the fat tail of the distribution.

There are two effects changing the wage distribution in this case - headhunters and skill-biased technological change. Skill-biased technological change increases wages of all workers and, therefore, moves the whole distribution to the right. To see the effect of only the headhunter channel on the wage distribution, Figure 1c plots the difference between the distributions without the effects of the skill-biased technological change. An interesting observation about the effect of headhunters on wage distribution can be done - the headhunter channel generates an effect similar to job polarization, namely, a decrease in the number of medium-paying jobs and an increase in the number of high- and low-paying jobs. This effect comes from the fact that low-skilled workers move from high-productive to low-productive firms (from the center to the left), and high-skilled workers move from low-productive firms to the high-productive firms (from the center to the right). The difference between the distributions also clearly indicates the appearance of a fatter right tail with the headhunter channel.

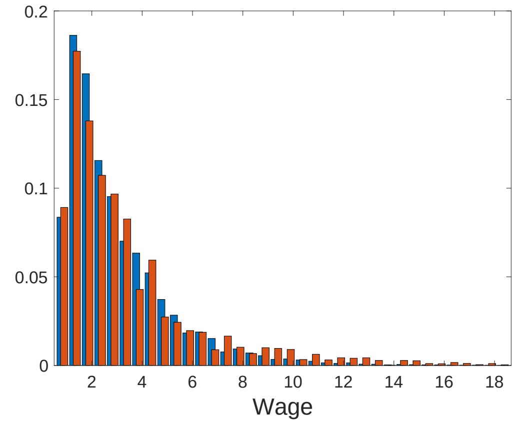
As it was stated before, the increase in wage inequality was mainly driven by the sharp increase in top wages. The top 1% wage share increased from 5.1% in 1970 to 10.9% in 2010 in the U.S., and the top 10% share increased from 25.7% to 34.5%. The top shares in 1970 were targeted in the calibration while the top shares in 2010 were not. The results of this experiment show how much of the overall increase in top wages can be explained by the additional channel in the labor market and an increase in the degree of complementarity in production. The results are presented in Table 2. In the model, the top 10% share increases by 5.1%, from 25.86% to 30.96%, while in the data it increases by 8.8%. The model is able to explain 58% of the actual increase in the top 10% wage share. For the top 1% wage share, the model predicts a 1.2% increase, while the actual increase is 5.8%. The model accounts for 21% of the actual increase in the top 1% wage share.

Table 2: Top Wage Shares in the Model and Data

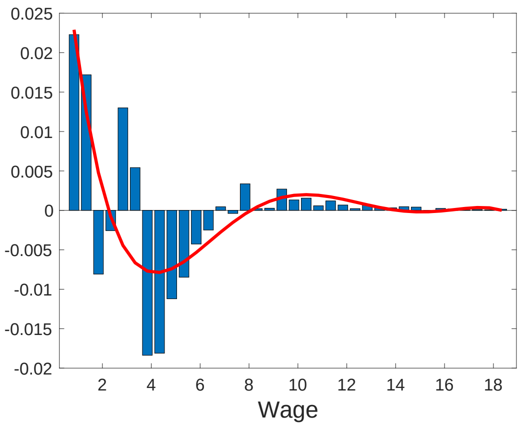
Model	Top 1%	Top 10%	Data	Top 1%	Top 10%
Without HH and SBTC	4.41%	25.86%	1970	5.1%	25.7%
With HH and SBTC	5.61%	30.96%	2010	10.9%	34.5%



(a) Distribution of Wages without the Headhunter Channel

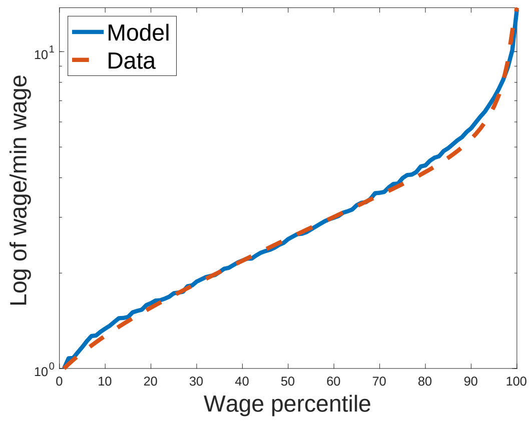


(b) Distribution of Wages with the Headhunter Channel

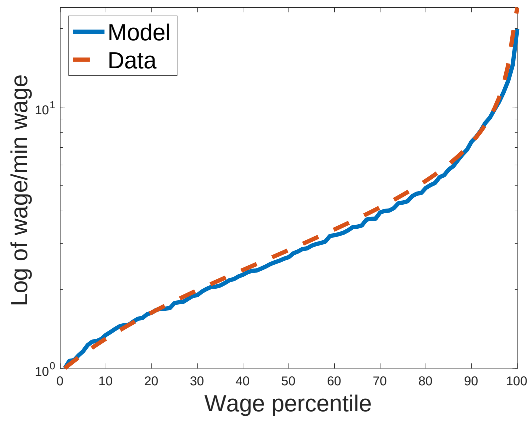


(c) Difference between the Distributions without SBTC

Figure 1: Distributions of Wages



(a) Distribution of Wages without the Headhunter Channel



(b) Distribution of Wages with the Headhunter Channel

Figure 2: Distributions of Wages in the Model and Data

3.2 Model Fit

The calibration of the model ensures that the model matches well the top shares of the wage distribution, but how does it match the overall wage distribution? To answer this question, Figure 2 plots the wage distributions in the model against the wage distributions in the data. The distributions in the data are from Song et al. (2019)[22]. I compare the model without the headhunter channel and skill-biased technological change to the data from 1981 and the model with the headhunter channel and skill-biased technological change to the data from 2013. The model matches well the overall shape of the wage distribution[23], especially at the top and the bottom, both in 1981 and in 2013. The model misses only slightly the middle of the wage distribution.

Another dimension of the data that can be used to assess the model fit is the change in within-group inequality. Lemieux (2006) shows that within-group inequality increased substantially for college graduates and post-graduates between the 1970s and the 2000s. There are no education groups in the model, so a direct comparison of the trends in the model and the data is not possible. However, there are well-defined skill types in the model with a wage distribution for each skill type. I compute within-group inequality in the model as the standard deviation of log wages by worker type following Uren and Virag (2011). Figure 3 plots the results for the model with and without the headhunter channel and skill-biased technological change. Within-group inequality decreases slightly for the

[22]Song et al. (2019) report the annual earnings rather than the wages, while in the model I use the wage rates. To bring the two as close as possible, I use the earnings distribution of workers with annual earnings of at least 2080 times the minimum wage. These workers are the most likely to have similar hours of work throughout the year, therefore, this distribution of earnings is the closest to the distribution of wage rates.

[23]The correlation between the model and the data is 0.9944 in 1981 and 0.9934 in 2013.

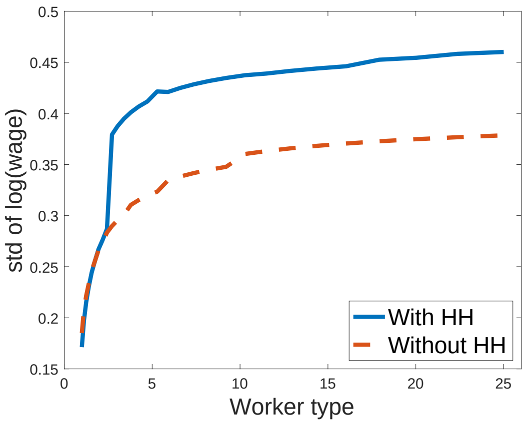


Figure 3: Within-group inequality by worker type

lowest-skilled workers and starts to increase for higher-skilled workers. The increase in within-group inequality is the largest for the workers above the headhunters' screening threshold. This happens because the headhunters make the distribution of wages of workers of such types bimodal (one peak for the workers hired through the standard channel and one peak for workers hired through the headhunter channel) substantially increasing the variance. The model matches qualitatively the substantial increase in within-group inequality for the highest-educated workers (who are more likely to be above the screening threshold) and a slight increase or decrease in within-group inequality for other groups.

3.3 Skill-Biased Technological Change

The large effect in Table 2 comes from the skill-biased technological change and the headhunter channel acting together. To assess the relative contributions of the headhunter channel and the skill-biased technological change to the increase of the top wages, I change separately only the matching technology (by adding the headhunter channel) or the degree of complementarity (SBTC). I also close the wage bargaining channel by fixing the wage structure to the one without the headhunters[24]. I present the results in Table 3. First, I fix the degree of complementarity on the level of 1970 but add the headhunter channel (bottom-left panel). In this case, the top 10% wage share is 28.66%, instead of 30.96% in the baseline calibration (upper-left), and the top 1% wage share is 4.99% (instead of 5.61%). The headhunter channel alone contributes to 32% of the increase in the top 10% wage share and 10% of the increase in the top 1% wage share in the data. To assess

[24] I use the wage function, $w(e,p)$, obtained in equilibrium without the headhunter channel. I apply this wage function to the joint distribution of workers and firms in equilibrium with the headhunter channel to obtain the counterfactual wage distribution. I use the same procedure to decompose the effect with and without SBTC using the wage function obtained without the headhunter channel but with and without SBTC, respectively.

how much of this increase is driven only by the improvements in matching, I close the wage bargaining channel (bottom-middle panel). Without wage bargaining, the top wage shares would rise to 27.5% and 4.61%. Therefore, improved matching alone explains 19% of the increase in the top 10% wage share and 3.5% of the increase in the top 1% wage share.

Table 3: Relative Contribution of Headhunters and SBTC

		HH	HH no WB	no HH
	Top 1%	5.61%	5.22 %	4.97%
SBTC	**Top 10%**	**30.96%**	**29.95%**	**27.84%**
	Δ90/50	0.53	0.57	0.21
	Δ50/10	0.11	0.03	0.11
	Top 1%	4.99%	4.61%	4.41%
no SBTC	**Top 10%**	**28.66%**	**27.50%**	**25.86%**
	Δ90/50	0.31	0.36	0
	Δ50/10	-0.04	-0.10	0

If, instead, I increase the degree of complementarity to the level of the baseline calibration without the headhunter channel (upper-right), the top 10% wage share increases to 27.84% and the top 1% wage share increases to 4.97%. The relative contribution of the degree of complementarity is about 23% (out of 58% in the baseline) for the top 10% wage share, and 10% (out of 21%) for the top 1% wage share. SBTC also explains a large fraction of the increase in the 90/50 and 50/10 wage ratios.

The interaction between SBTC and the headhunter channel is also important. The interaction explains around 3% of the increase in the top 10% wage share (58%-32%-23%) and 1% of the increase in the top 1% wage share (21%-10%-10%). With a higher degree of complementarity the importance of having a better match increases. Relative productivity of a firm with a high-skilled worker is even higher with respect to a similar firm with a low-skilled worker in case of a high degree of complementarity. Better assortative matching reinforces the effects of SBTC.

To give a chance to SBTC to explain a higher proportion of the rise in top shares, I recalibrate the SBTC to match the increase in the 90/50 wage ratio without the headhunter channel. I present the results in Table 4. We can see that the degree of complementarity must increase up to 1.23 without the headhunter channel to match the increase in the 90/50 wage ratio. If I match the 90/50 wage ratio, the model without headhunters still explains a smaller increase in the top 10%, but a slightly larger increase in top 1% wage share. The effect of SBTC on the 50/10 wage ratio is larger than the change in the data. The reason for this is that the rise of the degree of complementarity alone raises all the wages and the rise must be large to match the increase in the 90/50 wage ratio. We can see it in Figure 4. With the headhunter channel, all the wages rise only slightly due to a higher degree of complementarity (movement of the curve) and the top wages rise more than that due to improvements in the assortative matching (movement along the curve). With the headhunter channel, the high-skilled workers move up with the curve and move

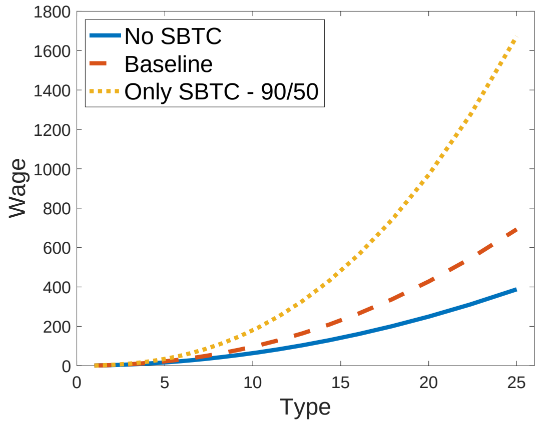


Figure 4: Wages of Workers in the Best-Fit Firms

along it to the right, and without the headhunter channel, they can only move up with the curve.

Table 4: Alternative Calibration of SBTC

	Data	Baseline	no HH, SBTC 90/50
	2010	$\gamma = 1.09$	$\gamma = 1.23$
Top 1%	10.9%	5.61%	5.73%
Top 10%	34.5%	30.96%	30.43%
$\Delta 90/50$	0.52	0.53	0.54
$\Delta 50/10$	0.16	0.11	0.25

These experiments show that skill-biased technological change helps to explain the rise in the 90/50 wage ratio that corresponds to the rise of the upper-middle class relative to the bottom but fails to replicate the sharp increase of the top wages[25]. The headhunter channel, instead, has the main effect on the top wages, rather than on the upper-middle class. Skill-biased technological change stretches the whole distribution to the right, while the headhunter channel fixes the left part of the distribution and moves the right tail further away.

3.3.1 Change in Wage Distribution

To demonstrate the lack of sufficient non-linearity of SBTC further, I compare the change in the wage distribution in different specifications of the model to the data. Figure IV in Song et al. (2019) shows the log change in the earnings distribution from 1981 to 2013 in

[25]These results are conditional on the functional form of SBTC determined by one parameter, γ. A more general functional form could lead to a greater increase in top wages explained by SBTC.

the U.S. The figure demonstrates that the earnings increased slightly for most of the distribution and exploded for top earners. The solid line on Figure 5 plots the corresponding change in log wages in the model with baseline calibration (with headhunters and SBTC). The shape of the line is very similar to the data (dashed line with asterisks) both at the bottom and at the top of the wage distribution. The skill-biased technological change contributes to the rise of the wages at the bottom and the middle of the distribution (see the dash-dotted line) and the headhunters generate the sharp increase in top wages (see the dotted line). It is evident that SBTC calibrated to match the increase in the 90/50 wage ratio (dashed line) is not able to reproduce the pattern observed in the data. Instead, it generates a very large increase in wages across all the distribution.

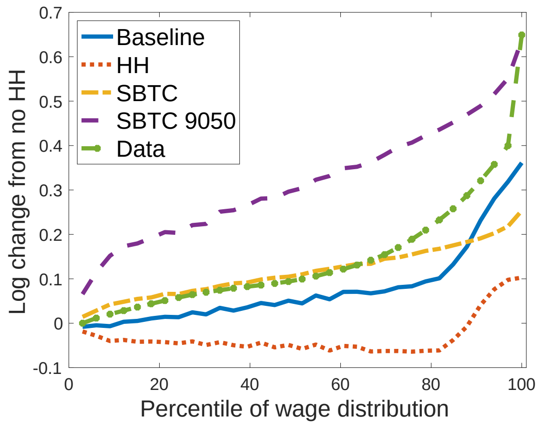


Figure 5: Wage growth by percentile of wage distribution, model-generated data, calibration comparison

Interestingly, the decomposition of the change into SBTC and headhunters is very similar to the decomposition into between- and within-firm changes in Song et al. (2019). It is not possible to decompose the change into between- and within-firm components in the model because each firm hires only one worker. With this caveat in mind, however, one could argue that the between-firm change in wages is driven mainly by SBTC with firms paying more initially being exposed to technological progress more. At the same time, the within-firm change might be driven by the change in the hiring practices, with the top-rank positions being filled through the headhunters.

3.4 Assortative Matching

The main mechanism behind the increase in wage inequality in the model is the increase in sorting between workers and firms, especially at the very top. With headhunters, high-skilled workers have an exclusive opportunity to be matched with high-productive firms, and high-productive firms, instead, have an exclusive opportunity to be matched with high-skilled workers. Empirically, there are two widely used ways to look at the assortative matching between workers and firms. First, one can directly compare the joint distributions of worker-firm matches over estimated types. And second, one can just look at the correlation between the types. I compute both statistics using the data simulated from the model in the baseline calibration in order to compare them to empirical estimates in the literature. The major drawback of this experiment, however, is that I can observe the real type of workers and firms directly, while in the data it is impossible.

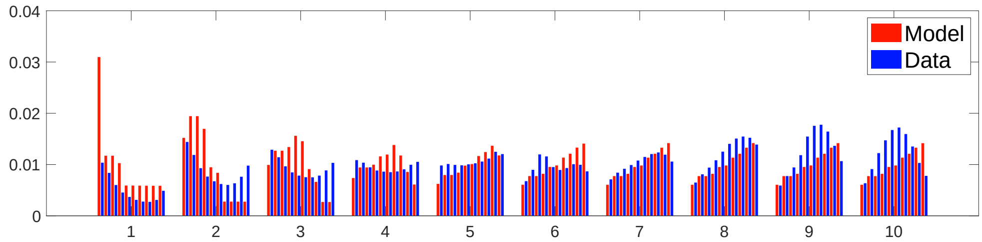
First, I study the change in the joint distribution of worker-firm matches. To do it, I split workers and firms into ten categories by their skill or productivity level and plot the joint distribution before and after introducing the headhunter channel. Red bars on Figure 6a show the distribution in the model without the headhunter channel, red bars on Figure 6b show the distribution with the headhunter channel, and red bars on Figure 6c show the change of the distribution. Numbers 1,2,3,..,10 in the figure correspond to the firm types, with 1 being the least productive firms and 10 being the most productive firms. For each number, one bar corresponds to one type of workers with the left-most bars being the lowest-skilled workers and the right-most bars being the highest-skilled workers.

As it can be seen from the figures, majority of high-skilled workers (within the top 20%) move to the best firms (top 20%). All other firms lose significantly in the share of top workers and gain in the share of lower-skilled workers. This pattern is strikingly similar to the findings of Song et al. (2019) who plot similar distributions for workers' and firms' fixed effects estimated on the U.S. data (blue bars on Figure 6). The model fits reasonably well the general patterns of joint distributions, especially in the middle of the distribution. The correlation between the model-generated distribution and the data in the 2000s is 0.70. The model delivers a reasonable fit of this very complicated object while only the change in the top two bars of the top two types of firms were targeted in the calibration of the model. Comparing the differences between the distributions in the model and the data (Figure 6c), the patterns are similar even for non-targeted types of firms, for example, the ones in the middle of the distribution.

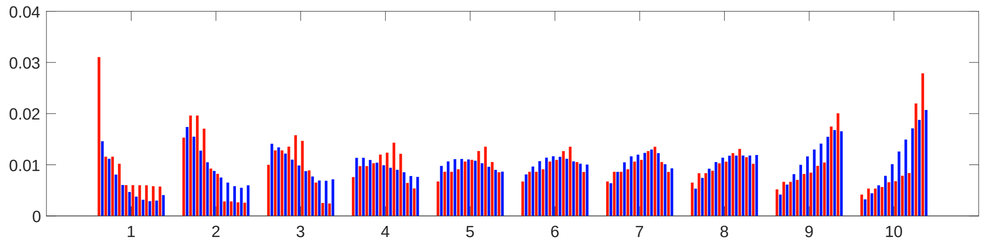
The second way to analyze sorting in the labor market is by computing correlations between the types of workers and firms. In order to do that, I draw 100,000 matches from the joint worker-firm distribution in the model and decompose the variance of log wages into worker type, firm type, the bargaining component, and the covariances between them. The left panel of Table 5 presents the results of this experiment for the steady-state without headhunters and SBTC, with headhunters and SBTC, and the difference between the two.

We can see that the covariance and the correlation of the worker and firm types increase significantly after the introduction of the headhunter channel and SBTC. Indeed, the increase is not only in the top part of the distribution but over the whole distribution. This happens because the high-skilled workers that move to the best firms free the positions for low- and medium-skilled workers in the rest of the firms also improving the matching for them.

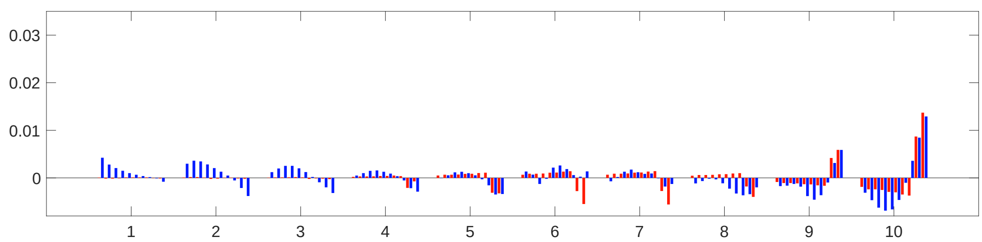
The right panel of Table 5 presents the results from Table V in Song et al. (2019), it shows the decomposition of the rise in earnings inequality between the 1980s and 2000s.



(a) Joint Distribution without the Headhunter Channel, Data in 1980s



(b) Joint Distribution with the Headhunter Channel, Data in 2000s



(c) Changes in the Joint Distributions

Figure 6: Joint Distributions of Worker-Firm Matches, Model and Data (Song et al. (2019))

Table 5: Wage Variance Decomposition, Model and Data

	Model				Data, Song et al. (2019)			
	No HH	HH	Δ	%	1980s	2000s	Δ	%
$Var(\log(w))$	0.316	0.418	0.102	-	0.762	0.903	0.141	-
$Var(\log(e))$	0.144	0.171	0.027	26.5	0.069	0.133	0.063	44.8
$Var(\log(p))$	0.185	0.222	0.037	36.3	0.068	0.074	0.006	4.3
$2Cov(\log(e), \log(p))$	0.086	0.140	0.054	52.9	0.036	0.096	0.058	41.2
$Var(\log(BP))$	0.031	0.034	0.003	2.9	-	-	-	-
$2Cov(\log(BP), \log(e))$	0.007	0.003	-0.004	-3.9	-	-	-	-
$2Cov(\log(BP), \log(p))$	-0.138	-0.152	-0.014	-13.7	-	-	-	-
Other components	-	-	-	-	0.589	0.601	0.012	8.5
$Cor(\log(e), \log(p))$	0.262	0.359	0.097	-	0.10	0.38	0.28	-

The overall change in wage inequality in the model is quite similar to the change in the data (0.102 as opposed to 0.141). Also, both in the model and in the data, a major part of the increase in the inequality is explained by the increase in the covariance of worker and firm types. In the model, 52.9% of the total increase in the variance of wages is explained by the increase in the covariance, in the data, the share is 41.2%. The model does well in allocating the relative importance of sorting and skill-biased technological change in the change of the variance of wages. Also, remarkably, the model with the headhunters matches well the correlation between worker and firm types observed in the data. The correlation is 0.36 in the model and 0.38 in the data and it was not targeted in the calibration.

The model does poorly in matching the roles of the increase in the variance of worker and firm types in the overall increase in the variance of wages. In the model, the variances of worker and firm types contribute similarly to the overall increase. In the data, instead, the increase in the variance of worker types is much more important relative to the variance of firm types. One possible explanation of this discrepancy is the difference in the definition of a firm in the model and in the data. In the data, a firm is a large entity with many employees while in the model a firm has only one employee and is closer to a position within a firm. If in the data the variance of firm types increases slightly but the variance of the positions within a firm increases substantially, the data estimation will likely interpret it as an increase in the variance of worker types. The model, instead, recognizing each position as a separate firm, will interpret it as an increase in the variance of firm types. Therefore, it is not surprising that the relative importance of firm and worker types is very different in the model and the data.

3.5 Alternative Decomposition - Headhunters as Profit-Maximizers

To assess relative contributions of headhunters and SBTC in the baseline model, I perform a decomposition of the total effect in Table 3. To determine the contribution of the headhunter channel to the increase in top wages, I compare the model calibrated to the 1970s to a model with the same technological parameters (no SBTC) but with the headhunter channel calibrated to the 2010s. This comparison, however, might be biased if the entry of headhunters was driven by the increase in top wages due to the increase in complementarity. The role of headhunters would be overstated in this case. The bias arises due to endogeneity of the headhunter entry.

To address the issue of endogeneity of the headhunter entry, I run an alternative decomposition of the effects of headhunters and SBTC on top wages. To do it, I extend the model to allow for the endogenous decision of headhunter entry. The headhunters need to collect vast data on firms and candidates to enter a new market (an industry or a type of position). Therefore, it is costly for the headhunters to increase the share of firms that might hire through the headhunters, χ. Assume that each headhunter optimally chooses the share to maximize the profit:

$$\max_{\chi} \left[FR(\chi) - C_{HH}(\chi) \right],$$

where $FR(\chi)$ is the total fee revenue of the headhunter and $C_{HH}(\chi)$ is the total screening cost. The fee revenues come from all the hires made through the headhunter in a given period:

$$FR(\chi) = \chi \int_{\hat{p}}^{\bar{p}} c_{fV}(p)\, dG(p, \overline{\chi}),$$

where $G(p, \overline{\chi})$ is the vacancy distribution given an average coverage by headhunters. Assume that the cost is quadratic in the coverage share:

$$C_{HH}(\chi) = \frac{c_{HH}}{2} \chi^2.$$

The first-order condition for the optimal headhunter coverage implies

$$\chi = \frac{1}{c_{HH}} \int_{\hat{p}}^{\bar{p}} c_{fV}(p)\, dG(p, \overline{\chi}).$$

I calibrate the screening cost, c_{HH}, to match the headhunter coverage in the baseline calibration with the headhunters and SBTC. Then, I use this cost to compute the optimal headhunter coverage in the equilibrium without SBTC. I then compute an alternative decomposition of the total effect of headhunters and SBTC on the rise of the top wages. The alternative decomposition is presented in Table 6. With a fixed coverage rate (second column), the headhunters explain 32% of the increase in the top 10% wage share and 10% of the increase in the top 1% wage share. However, for this coverage rate to be sustainable in the equilibrium with headhunter entry, the screening cost must have been much lower in the 1970s (0.95 instead of 1.17) that is clearly not realistic. If the screening costs were the same, the free entry of headhunters would lead to a maximum coverage rate of 26%, instead of 30% in the baseline. This coverage rate is significantly lower but it has only a minor effect on the top wage shares. With optimal entry, the headhunters would still

Table 6: Alternative Decomposition

	Baseline	no SBTC	no SBTC, entry	no SBTC, no HH
Share of top firms	$\chi = 0.3$	$\chi = 0.3$	$\chi = 0.257$	$\chi = 0$
Top 1%	5.61%	4.99%	4.82%	4.41%
Top 10%	30.96%	28.66%	28.31%	25.86%
Screening cost, c_{HH}	1.17	0.95	1.18	-

explain 28% of the increase in the top 10% wage share and 7% of the increase in the top 1% wage share.

The effect of the endogenous entry on the explained increase in the top wage shares can be viewed through the elasticity of the top shares with respect to the coverage rate. As can be seen in the robustness exercises in the Appendix, the elasticity is rather small for small coverage rates, so even a 13% decrease in the coverage rate doesn't lead to a large drop in the top shares. The reason for this might be the following. Even for a low coverage rate, say 5% per period, the top shares in a steady state will be populated mostly by workers hired through the headhunter channel. Increasing the coverage ratio further does not increase the number of top workers in top firms much but significantly decreases the number of top workers in low-productivity firms. Therefore, a further increase in the coverage rate compresses the bottom of the distribution rather than expanding the top.

This alternative decomposition demonstrates that the rise of headhunters explains the majority of the increase in top wages even after accounting for the endogenous entry of headhunters. Importantly, it also suggests that to match the limited role of headhunters observed in the 1970s, the model would require higher screening costs than in the 2010s. These results are supportive of the idea that headhunters increased their market share because technological advancements decreased their screening costs, rather than because of rising top wages. While both effects must be present in practice, this decomposition suggests that screening costs play a more important role.

4 Headhunter Industry

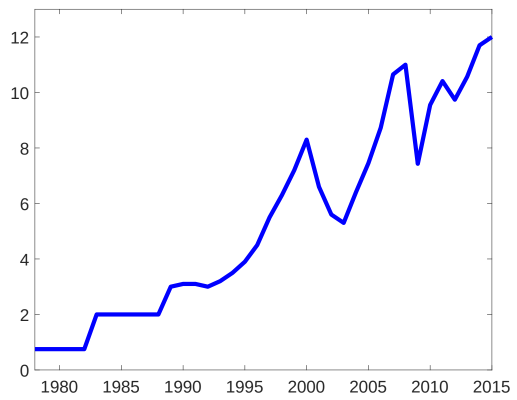
Individual headhunters are typically focused on a specific position or industry and collect detailed databases with information on the majority of potential candidates for such position or industry. With this detailed information already collected, when asked to assist to fill a position, they can choose the best fitting candidate and improve matching. As the headhunters already possess the information on the majority of candidates, they are more efficient in screening than firms. Firms could carry out the selection without the help of a headhunter but they would have to pay the screening cost, that can be immensely high, just to hire one candidate (say a CEO). Headhunters, using the same database to place candidates in different companies, spread the screening cost across many hires, therefore improving efficiency. After a headhunter chooses a candidate from its database, it calls the candidate directly and asks whether she wants to consider a job offer (without specifying the offer). The headhunter contacts any candidate who is perceived to be the best fit for the position without the candidate having to signal interest in changing her job. A worker who has not put effort into receiving an offer from a headhunter and who agrees to consider the offer is essentially searching passively on the job[26]. Passive search helps high-skilled workers not to get stuck for long in positions that do not fit them, moving them to a better fitting position and improving matching.

The main question is how big is the headhunter industry. To determine the exact market share of a very closed and private industry is a difficult task. Few headhunter companies release the number of hires in a given year. To overcome the shortage of data, one can use IACPR report (2003) that claims that 54% of the positions above $150,000 a year were filled by headhunters in 2003. Another way to determine the market share of the headhunters is to compare the estimates of the fee revenues to the ones implied by the total wage bill. Total fee revenues in the U.S. are around $4.6 billion as estimated by the Association of Executive Search and Leadership Consultants (AESC). It is possible to compute the total wages that go to the top 5% of the U.S. employees using the top 5% wage share. Then, using the hiring rate, one can determine total wages that go to the new hires in a given year. Headhunters receive a fee of around 30% of the first year wage paid to the new hires. It is possible to determine what would be the aggregate fee revenues for a given market share of headhunters. Given the average hiring rate of 3.5%, the share of headhunters in the labor market for positions in the top 5% must be around 15%, to be consistent with the estimates by AESC. However, the hiring rate at the top is, in general, much lower than in the lower-paying jobs, with tenures being significantly longer. With a more realistic hiring rate, the implied market share of headhunters is more than 30% but still below 54% estimated by IACPR.

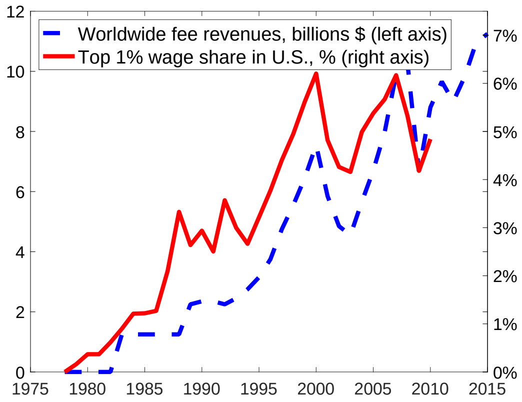
The history of the rise of headhunters started in the U.S. already in the 1950s. However, the first decades were not very successful for them. Only in the late 1970s and early 1980s, the industry started to expand sharply with worldwide fee revenues rising from $0.75 billion in 1978 to $3.9 billion in 1990. The fee revenues kept rising up to $12 billion in 2015 (Figure 7a). This rise was partly mechanical because top wages were also rising over the same period (Figure 7b), but the revenues increase was much larger in proportion to the increase of the wages. Another indicator of the expansion of the industry is the number of hires by headhunters. For example, the number of assignments of one of the historical leaders of the industry Korn/Ferry increased from 42 in 1969 to 8,480 in 2015 according to the financial statements.

There might be several reasons for the rise of headhunters. The most important reason is the technological progress in IT and communication that increased the quality of the services provided by headhunters. Development and growing availability of computers reduced the costs of managing and searching through databases of potential candidates. Communication technology (mobile phones and emails), instead, made it easier to contact potential candidates and allowed headhunters to expand their networks of potential

[26] Cappelli and Hamori (2013) show that more than half of executives are willing to consider an offer when a headhunter calls them.



(a) Estimated Worldwide Fee Revenues of Headhunters



(b) Cumulative Change in Headhunters Fee Revenues and the Top 1% Wage Share

Figure 7: Estimated Worldwide Fee Revenues of Headhunters and Top Wages, Calculated from AESC and Piketty (2014)

candidates. Companies that adopted new technologies earlier were more successful[27]. Another effect of technology goes through the demand side: better technology made it easier to apply for jobs (especially in the late 1990s and 2000s). More applications increased the amount of information that the firms had to evaluate to hire a worker, and the higher was the position, the more information was there to evaluate. It became more efficient for the firms to delegate the screening of applicants for top positions to intermediaries - the headhunters - and the demand for headhunter service increased. One more potential reason for the rise that goes through the demand side is related to the nature of the skills required from employees in the top positions. Because of technological change, globalization, or change of company structure, it became more important for the firm to hire employees with higher general skill in comparison to the 1970s. Firms started to use headhunters more because in the 1990s the skill of the CEO, for example, affected the performance of the company much more than in the 1970s. Even though there is more evidence in support of the supply story, this paper doesn't exclude other reasons for the rise of headhunters.

[27] Jenn (1995) writes: "The drive towards a more consistent quality of service throughout the world has been greatly assisted by the application of information technology to the search business and the use of global databases. Technological advances have allowed firms to search more widely and communicate more efficiently. Virtually all executive search firms are attempting to modernise their communications and database systems on a global basis. ... This is the area where the search world is changing most dramatically. Firms have a tremendous opportunity to improve their efficiency, achieve better margins and differentiate themselves from their competitors."

5 Suggestive Evidence

5.1 Micro Evidence

This section presents the empirical analysis of the potential effect of headhunters on the CEO compensation. CEOs constitute a major part of the hires by headhunters, accounting for 20000 hires by headhunters in 2013 in the U.S. alone, and therefore are a good proxy for individual effects of headhunters on the matching between workers and firms.

5.1.1 Data and Estimation

The data on CEO compensation and the firm level data are obtained from COMPUSTAT dataset. In particular, following Gabaix, Landier and Sauvagnat (2014), the variable TDC1 of EXECUCOMP panel is used to measure CEO compensation. TDC1 includes salary, bonus, restricted stock granted and the Black-Scholes value of stock options granted. Also following Gabaix, Landier and Sauvagnat (2014), four proxies for the firm size will be used: firm value, equity value, sales, and income. All four proxies are constructed from variables obtained from COMPUSTAT yearly dataset[28]. Industry dummies are constructed using the four-digit SIC industry codes as in Fama and French (1997). A dummy variable for a change of the CEO is constructed such that it is equal to 0 if the CEO is the same as the CEO of the first observation of the company, and 1 otherwise:

$$NewCEO_{i,t} = \begin{cases} 1 & \text{if CEO is different from the first observation of the firm} \\ 0 & \text{otherwise.} \end{cases}$$

Another important variable that will be analyzed is the index of enforceability of non-competition constructed by Garmaise (2011). The index is higher in the states where the non-compete agreements are enforced by courts and low in the states were the non-compete agreements are forbidden. Non-compete agreements restrict job-to-job transitions for workers and therefore limit the activity of headhunters.

The following sample will be used. The time period analyzed is from 1993 to 2013. The analysis will be restricted only to the CEO of every U.S. based company in the dataset. If a firm changes the CEO more than once during the sample period, all observations starting from the third CEO are dropped. These restrictions leave 3102 firms with 7.95 average years of observation.

I estimate the following equation:

$$\log(TDC1_{i,t}) = \alpha * NewCEO_{i,t} + \beta * \log(Firm\ size_{i,t}) + FE_t + FE_i + \varepsilon_{i,t},$$

where $TDC1_{i,t}$ is the CEO compensation in firm i and year t, $NewCEO_{i,t}$ is the dummy variable constructed as described above, and $Firm\ size_{i,t}$ is one of the four measures of the firm size of firm i and year t.

[28] Detailed description is in the appendix

5.1.2 Results

Table 7 presents the results of the estimation using the full sample as well as two sub-samples when headhunter fee revenues were increasing particularly fast (as seen from Figure 7a) - 1993 to 1998 and 2004 to 2007. Columns (1) and (2) present the results of the estimation with firm fixed effects and with or without the year fixed effects for the full sample. The results show that after a company changes the CEO it pays her from 5% to 16% more per year than to the previous CEO controlling for the firm size. The effect is even stronger if we focus on two sub-samples with the fast industry growth. The coefficient increases from 5% to 9% in the first sub-period (column (3)) and from 5% to 13,6% in the second sub-period (column (4)). This can be viewed as an indirect evidence of a higher use of headhunters during those periods and, therefore, better improvements in the matching between CEOs and firms resulting in higher compensation.

Table 7: CEO Compensation and the Change of the CEO

	Full Sample		1993-1998	2004-2007
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.1579	0.0495	0.0906	0.1364
	(0.0296)	(0.0180)	(0.0315)	(0.0382)
Firm Size Controls	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R^2	0.66	0.66	0.724	0.785
Number of observations	24673	24673	8304	4505

The important question is what is the channel of this effect, why are firms paying more to new CEOs? One potential explanation can be that the new CEO has a higher bargaining power than the previous CEO, it can be the case especially if the new CEO was hired with the help of a headhunter while the previous CEO was not. To test this channel, I augment the estimated equation with the interaction term between the $NewCEO$ dummy variable and the measures of the firm size. The coefficient of the measure of the firm size in this regression can be viewed as the bargaining power of the CEO, i.e. how much his compensation increases when the firm is growing, and the interaction term can be viewed as the change in the bargaining power between the new and the previous CEOs. The results (presented in the Appendix) show that the interaction term is negative or not significantly different from zero. These results suggest that the increase of CEO compensation does not come from a higher bargaining power.

In the model, matches are better with headhunters, productivity is higher and drives the wages up. To explore the matching channel, I add the non-compete enforceability index to the analysis. In Table 8 I present the results of the estimation where I add the interaction between the non-compete enforceability index and the new CEO dummy. The interaction term increases the magnitude of the effect of the new CEO dummy and has a negative and significant coefficient by itself. It means that the effect of the change of the

CEO on the compensation is higher in the states with low non-compete enforceability index and it decreases with a higher index. It is also interesting to notice that in the states where the index would be 1 (the highest index is 0.9 in Florida) the overall effect of the change of the CEO on the compensation would be negative.

Table 8: CEO Compensation, New CEOs, and the Non-Compete Enforceability Index

	Sample period 1993 - 2013		
Log of compensation	(1)	(2)	(3)
NCEI*New CEO	-0.1685	-0.2377	-0.1805
	(0.0475)	(0.0477)	(0.0471)
New CEO	0.1245	0.0778	0.0663
	(0.0282)	(0.0190)	(0.0191)
Firm Size Controls	Yes	Yes	Yes
Year FE	No	Yes	Yes
Industry FE	Yes	No	Yes
R^2	0.409	0.381	0.419
Number of observations	24217	24217	24217

5.1.3 Discussion

The empirical results presented in this section suggest that the matches between CEOs and companies improve over time in the U.S. The improvement in the matching results in higher CEO compensation and larger size of the firms. The fact that the matching is improving over time supports the mechanism discussed in the paper. Headhunters provide better matches for firms and CEOs increasing the firm size and the CEO compensation.

Of course, there are shortcomings in this empirical specification because we don't know which CEOs are hired through a headhunter and which CEOs come from internal promotion or other channels. To address this issue directly, one needs to collect the data on the origin of the CEO and the way she was hired. Such study would be able to analyze the difference between CEO compensation for a CEO coming through headhunters and not. Most importantly, it would be also able to determine the effect of a CEO hired by a headhunter on the firm performance. However, such datasets are not available at this moment.

Closest to that idea, Murphy and Zabojnik (2007) provide an empirical evidence on the CEO origins at the moment of her appointment, i.e. whether she is coming from within the company or from outside, and the effect of the origin on the compensation. They study the S&P 500 companies during the period from 1970 to 2005. They show that during the 1970s and the 1980s only 15% and 17% of CEO appointments account for the outside hires, while it increased to 26% in the 1990s and almost 32.7% in the 2000s. Even more importantly, they show that external CEOs receive 14.2% higher compensation on average over the full sample, with the difference being just 6% in the 1970s, 15,9% in the 1980s and 19.6% in the 1990s. Not only the companies rely more and more on the outside CEOs but also the pay difference between the internal and external CEOs is increasing. Among other possible explanations, the results by Murphy and Zabojnik (2007) are consistent with the rise of the headhunters over the last 40 years and their improved efficiency over time.

To try to overcome the lack of data on identities of the CEOs hired by headhunters, I use non-compete enforceability index as a proxy for the probability to be hired by a headhunter. In the states with a high NCEI, activity of headhunters is limited and, therefore, very few positions are filled by headhunters. The results show, indeed, that in the states with low NCEI the increase in CEO compensation after a CEO change is larger. This suggests that more CEOs are hired by headhunters in the states with low NCEI, so the improvement in matching is stronger and it leads to higher compensation.

Would these results be consistent with alternative explanations of the increase in CEO compensation? Two of the alternative explanations are the skill-biased technological change and changes in wage bargaining. Both explanations would lead to higher compensation of CEOs in general. However, under these alternative explanations, the compensation would rise for all CEOs, both incumbents and newly hired. That is, there would be no reason for the compensation of the new CEO to be systematically higher than the compensation of the previous CEO after controlling for the general trend in compensation and firm size. In fact, the coefficient of the new CEO dummy is lower in regressions with year fixed effects (columns (1) and (2) of Table 7) suggesting that there is a general trend in CEO compensation. This general trend raises the compensation of all CEOs and might be explained by the skill-biased technological change or changes in general wage bargaining. The fact that the coefficient of the new CEO dummy is positive suggests that the compensation is higher only for new CEOs, so a purely technological story can be excluded[29]. It must be the case that the new CEOs are better than previous ones or have a better bargaining position. If new CEOs are better, the matching is likely improving. If instead, they have a better bargaining position, there must be an underlying quality of the CEOs that makes their outside options better than the ones of the previous CEOs, signaling that the quality of the new CEOs is better at least in some dimension[30]. Therefore, the results presented in this section together with the results by Murphy and Zabojnik (2007) are consistent with the mechanism proposed in this paper. Although the evidence is not definitive as there might be other explanations consistent with both results.

Other studies discussing the increase in CEO compensation over the past decades offer various explanations of this phenomena. Gabaix and Landier (2008) and Gabaix, Landier and Sauvagnat (2014) argue that the CEO pay increases because the average company size is increasing. Murphy and Sandino (2010) argue that the CEOs may extract a larger

[29]Alternatively, the CEOs might have fixed-wage long-term contracts and can rebargain the wage only with a new employer. This could explain the positive coefficient of the new CEO dummy with skill-biased technical change. However, this story seems to be unrealistic as the majority of CEOs have pay-for-performance contracts and their bonuses are decided annually by the boards of directors.

[30]Moreover, additional empirical evidence presented in the Appendix suggests that the bargaining power of the new CEOs is not higher than the bargaining power of previous CEOs. Hence, if the bargaining power increased, it increased for all CEOs, both new and previous.

rent from the company by hiring external compensation consultants that follow their interest. Murphy and Zabojnik (2004) show that the nature of CEO skills required to successfully run a company is changing over time and, therefore, more firms hire the CEOs from outside of the firm and have to pay her more.

Among other studies closely related to the mechanism studied in this paper, Garmaise (2011) shows that tougher non-compete agreements regulation reduces CEO turnover and compensation. Again, this suggests that the activity of headhunters is limited in the states with higher NCEI[31] and, therefore, it reduces opportunities for CEOs to transit between firms and improve the efficiency of matching limiting compensation. Pan (2017) shows the importance of assortative matching between CEOs and firms for determination of the CEO compensation and the firm's performance. However, Pan (2017) doesn't consider the change in the degree of assortative matching over time or geographical differences.

5.2 Cross-Country Comparison

Headhunters entered labor markets of different countries in different periods and therefore were used by firms to a different extent[32]. This variation in headhunters' activity allows me to argue about the causality between the role of headhunters and the growth of top wages. This section uses the data on the major headhunter companies in European labor markets in 1997 and the data on top income shares between 1980 and 2010. Ideally, top wage shares should be used in the analysis, however, such data is not available for all countries over the whole period in question.

Data on major headhunter companies operating in Europe in 1997 is available in Jenn (1999). The data includes the distribution of fee revenues as well as the number of hires by country[33]. I aggregate the data by country to get total fees and a total number of hires by headhunters in a country in 1997. I normalize the data by GDP in 1997 (for fee revenues) or total employment in 1997 (for hires). The normalization allows comparing the share of headhunters between countries. The question that this analysis answers is what is the relation between the headhunter activity and the dynamics of top incomes? To answer this question, Figure 8[34] plots the relations between normalized hires by headhunters and the top 10% income share, or growth of the top 10% income share. Similar relations for the top 1% income share as well as for normalized fee revenues are presented in the Appendix. Figure 8a shows that there is a strong positive correlation between normalized hires by headhunters and the future growth of the top 10% income share. Only Norway falls from the general pattern. Norway experienced a change in capital income taxation in 2006, so most likely this drop is not related to labor incomes.

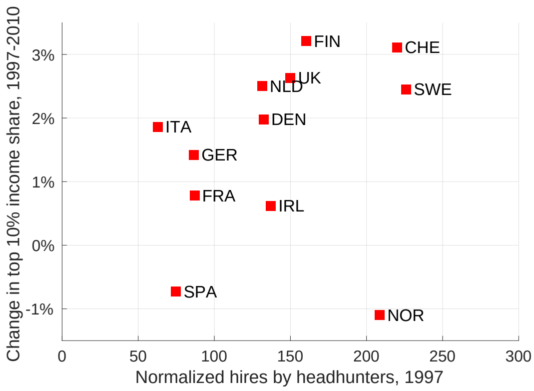
To address the concern that headhunters were more active in countries where top wages were already higher, Figure 8b plots the relation between top income shares in 1997

[31] Garmaise (2011) does not talk about headhunters in his study.

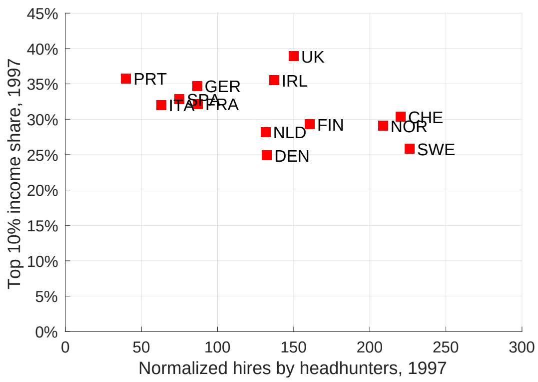
[32] For example, due to different labor market legislation, language barriers and other institutional reasons.

[33] The number of hires is estimated and not exactly observed for some companies.

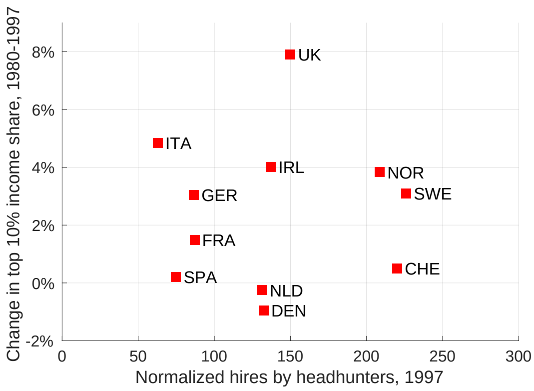
[34] Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.



(a) Top 10% Income Share Growth, 1997-2010, and Normalized Hires by Headhunters, 1997



(b) Top 10% Income Share, 1997, and Normalized Hires by Headhunters, 1997



(c) Top 10% Income Share Growth, 1980-1997, and Normalized Hires by Headhunters, 1997

Figure 8: Top 10% Income Share and Normalized Hires by Headhunters, Calculated from Jenn (1999), OECD, and WID

and normalized hires by headhunters in 1997. As it is evident from the figure, there is no correlation between headhunter activity and top income shares in 1997. It means that differences of headhunter intensity across countries are driven by other factors, exogenous to top income levels. To further strengthen this claim, Figure 8c plots the growth of top income shares before 1997 against normalized hires in 1997. Lack of positive correlation shows that headhunters intensity is not driven by the previous growth of top incomes. Headhunters didn't choose countries with fast-growing top incomes.

All the results for normalized hires hold also for normalized fee revenues. This analysis suggests two important facts. First, it suggests that headhunter activity, indeed, signals the future growth of top incomes. Second, this analysis suggests that the distribution of headhunter activity over countries is exogenous to the level of top incomes and the history of the growth of top incomes. There must be other factors limiting headhunter activity in some countries, for example, labor market legislation or higher costs of establishing detailed databases of potential candidates. This evidence, however, does not provide any hints on the mechanism of the top incomes increase, or the degree of the quality of the matching.

The importance of this empirical evidence is in demonstrating the lack of reverse causality. Headhunters might be more active in countries where the income inequality was higher so they came to the markets to extract higher fee revenues. In this case, the increasing top wages would drive the rise in the headhunter industry, and the mechanism presented in this paper would be weaker. However, the results presented in this section suggest that only the future change in top incomes is correlated with the headhunter intensity, so reverse causality can be rejected.

6 Conclusion

This paper introduces the headhunter channel to the standard model of random matching. The fact that headhunters have better information about a worker's skill level and that they can approach workers who are not actively searching for a (new) job at this moment allows for better screening of workers and reduces labor market frictions at the top part of the wage distribution. Moreover, headhunters separate the labor market for high and low-productive firms allowing the high-productive firms to access only the high-skilled workers. Because of worker skill and firm productivity complementarities, the wages of workers hired through headhunters increase more than proportionally to the rest of the workers. Thus, the presence of headhunters generates a fat tail of the wage distribution with a larger wage share of the top 1% and 10% workers.

Quantitative analysis shows that introduction of the headhunter channel in otherwise standard random matching model accounts for 32% of the increase in the top 10% share of wages and 10% of the increase of the top 1% share of wages in the U.S. between 1970 and 2010. The main effect comes from the improvement in the assortative matching between workers and firms, especially at the top. The pattern and the amplitude of the improvement are comparable to the empirical estimates of the change in assortative matching in the U.S. over the same period. The headhunter channel helps to generate the strong non-linearity in the pattern of matching observed in the data.

The paper also shows that the new CEOs in the U.S. get higher compensations comparing to the previous CEOs in the same companies and this effect is weaker in the states with high non-compete enforceability index, i.e., in the states that potentially limit the activity of headhunters. Then, the paper also provides the empirical evidence of the joint increase of the use of headhunters by firms and the top income shares. The paper uses cross-country data on headhunter revenues and number of hires through headhunters together with the top income shares to show that normalized hires by headhunters are a good predictor of the future growth of the top income shares in European countries.

References

Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, 40(1): 7–72.

AESC. 2015. "Insights, Q4 2015, State of the Executive Search Industry."

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. 2013. "The Top 1 Percent in International and Historical Perspective." *Journal of Economic Perspectives*, 27(3): 3–20.

Aoki, Shuhei, and Makoto Nirei. 2017. "Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the United States." *American Economic Journal: Macroeconomics*, 9(3): 36–71.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the US Labor Market." *American Economic Review*, 96(2): 189–194.

Bagger, Jesper, and Rasmus Lentz. 2019. "An Empirical Model of Wage Dispersion with Sorting." *The Review of Economic Studies*, 86(1): 153–190.

Bagger, Jesper, Kenneth L Sørensen, and Rune Vejlin. 2013. "Wage Sorting Trends." *Economics Letters*, 118(1): 63–67.

Bell, Brian D., and John Van Reenen. 2013. "Extreme Wage Inequality: Pay at the Very Top." *American Economic Review Papers and Proceedings*, 103(3): 153–157.

Borovičková, Katarína, and Robert Shimer. 2017. "High Wage Workers Work for High Wage Firms." NBER Working Paper No. w24074.

Cappelli, Peter, and Monika Hamori. 2013. "Who Says Yes When the Headhunter Calls? Understanding Executive Job Search Behavior." NBER Working Paper No. w19295.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying Sorting - in Theory." *Review of Economic Studies*, 78: 872–906.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2017. "Job Search Behavior among the Employed and Non-Employed." NBER Working Paper No. w23731.

Fama, Eugene F., and Kenneth R. French. 1997. "Industry Costs of Equity." *Journal of Financial Economics*, 43(2): 153–193.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123(1): 49–100.

Gabaix, Xavier, Augustin Landier, and Julien Sauvagnat. 2014. "CEO Pay and Firm Size: an Update after the Crisis." *Economic Journal*, 124(574): F40–F59.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica*, 84(6): 2071–2111.

Garmaise, Mark J. 2011. "Ties that Truly Bind: Non-competition Agreements, Executive Compensation and Firm Investment." *Journal of Law, Economics, and Organization*, 27(2): 376–425.

Gautier, Pieter A., and Coen N. Teulings. 2015. "Sorting and the Output Loss due to Search Frictions." *Journal of the European Economic Association*, 13(6): 1136–1166.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. "Identifying Equilibrium Models of Labor Market Sorting." *Econometrica*, 85(1): 29–65.

IACPR. 2003. "International Association of Corporate and Professional Recruiters Report."

Jenn, Nancy Garrison. 1995. *Executive Search in the Americas: Choosing and Using a Headhunter.* Economist Intelligence Unit.

Jenn, Nancy Garrison. 1999. *Executive Search in Europe: Choosing and Using a Headhunter.* Economist Intelligence Unit.

Jones, Charles I. 2015. "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality." *Journal of Economic Perspectives*, 29(1): 29–46.

Jones, Charles I, and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy*, 126(5): 1785–1826.

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill." *American Economic Review*, 96(3): 461–498.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent. 2009. "Performance Pay and Wage Inequality." *Quarterly Journal of Economics*, 124(1): 1–49.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2016. "Matching, Sorting and Wages." *Review of Economic Dynamics*, 19: 63–87.

Manning, Alan. 2011. "Imperfect Competition in the Labor Market." In *Handbook of Labor Economics*. Vol. 4, 973–1041. Elsevier.

Murphy, Kevin J., and Jan Zabojnik. 2004. "CEO Pay and Appointments: A Market-Based Explanation for Recent Trends." *American Economic Review*, 94(2): 192–196.

Murphy, Kevin J., and Jan Zabojnik. 2007. "Managerial Capital and the Market for CEOs."

Murphy, Kevin J., and Tatiana Sandino. 2010. "Executive Pay and "Independent" Compensation Consultants." *Journal of Accounting and Economics*, 49(3): 247–262.

Pan, Yihui. 2017. "The Determinants and Impact of Executive-Firm Matches." *Management Science*, 63(1): 185–200.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Harvard University Press.

Pissarides, Christopher A. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages." *American Economic Review*, 75(4): 676–690.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica*, 70(6): 2295–2350.

Schulz, Bastian, and Benjamin Lochner. 2016. "Labor Market Sorting in Germany." CESifo Working Paper No. 6066.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95(1): 25–49.

Shimer, Robert, and Lones Smith. 2000. "Assortative Matching and Search." *Econometrica*, 68(2): 343–369.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2019. "Firming Up Inequality." *The Quarterly Journal of Economics*, 134(1): 1–50.

Teulings, Coen N., and Pieter A. Gautier. 2004. "The Right Man for the Job." *Review of Economic Studies*, 71(2): 553–580.

Uren, Lawrence, and Gabor Virag. 2011. "Skill Requirements, Search Frictions and Wage Inequality." *International Economic Review*, 52(2): 379–406.

A Appendix

A.1 Model Equations

Workers

The value of search through both channels for a high-skilled unemployed worker is:

$$
\begin{aligned}
S_{UVH}(e) \equiv \; & f_H(\theta_H)\left(\int_{\hat{p}}^{\overline{p}} (W(e,p) - U(e))\, dG(p) - c_{wH}(e)\right) \\
& + f_V(\theta_V)(1 - f_H(\theta_H)) \\
& \cdot \int_{\underline{p}}^{\hat{p}} (W(e,p) - U(e))\, dG(p) \\
& - c_{wV}(e).
\end{aligned}
$$

The value of search through both channel for a high-skilled employed worker is defined as:

$$
\begin{aligned}
S_{EVH}(e,p) \equiv \; & f_H(\theta_H)\left(\int_{\hat{p}}^{\overline{p}} (W(e,p') - W(e,p))\, dG(p') - c_{wH}(e)\right) \\
& + f_V(\theta_V)(1 - f_H(\theta_H)) \\
& \cdot \int_{p}^{\hat{p}} (W(e,p') - W(e,p))\, dG(p') \\
& - c_{wV}(e),
\end{aligned}
$$

note that in this case the first integral starts always in $\hat{p}$ because it will never be optimal to search through both channels if a worker is already working in a firm that hires through the headhunter channel.

Firms

The value of hiring through the vacancy channel in this case is:

$$
\begin{aligned}
V_V(p) = \; & -c_{fV}(p) + \beta\Bigg(V(p) + q_V(\theta_V)\Bigg(\frac{u_V}{u_V + a_V}\int_{\underline{e}}^{\hat{e}} (J(p,e) - V(p))\, dU(e) \\
& + \frac{u_V}{u_V + a_V}(1 - f_H(\theta_H))\int_{\hat{e}}^{\overline{e}} (J(p,e) - V(p))\, dU(e) \\
& + \frac{a_V}{u_V + a_V}\int_{\underline{e}}^{\hat{e}} \frac{\Lambda_V(e,p)}{\Lambda_V(e,\overline{p})} (J(p,e) - V(p))\, dL_V(e) \\
& + \frac{a_V}{u_V + a_V}(1 - f_H(\theta_H))\int_{\hat{e}}^{\overline{e}} \frac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})} (J(p,e) - V(p))\, dL_{VH}(e)\Bigg)\Bigg),
\end{aligned}
$$

where the first part in the summation is the expected value of a match after meeting a low-skilled unemployed worker, the second - a high-skilled unemployed worker, the third - a low-skilled employed worker, and the forth - a high-skilled employed worker.

Similarly, the value of hiring through the headhunter channel is:

$$
\begin{aligned}
V_H(p) = \; & -c_{fH}(p) + \beta\Bigg(V(p) + q_H(\theta_H)\Bigg(\frac{u_H}{u_H + a_H}\int_{\hat{e}}^{\overline{e}} (J(p,e) - V(p))\, dU(e) \\
& + \frac{a_H}{u_H + a_H}\int_{\hat{e}}^{\overline{e}} \frac{\Lambda_H(e,p)}{\Lambda_H(e,\overline{p})} (J(p,e) - V(p))\, dL_H(e) \\
& + \frac{a_H}{u_H + a_H}\int_{\hat{e}}^{\overline{e}} \frac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})} (J(p,e) - V(p))\, dL_{VH}(e)\Bigg)\Bigg),
\end{aligned}
$$

where the first part in the summation is the expected value of a match after meeting a high-skilled unemployed worker, the second - a high-skilled employed worker searching only through the headhunter channel, and the third - a high-skilled employed worker searching through both channels.

Now, given the distributions, we can also specify the quit rate of a worker with skill e from a firm with productivity p:

$$s_Q(e,p,\omega)=\begin{cases} f_V(\theta_V)\left(\frac{G(\hat{p})-G(p)}{G(\hat{p})-G(\underline{p})}\right) & if\ p<\tilde{p}_V(e)\ \ and\ e<\underline{e} \\ f_H(\theta_H)\left(\frac{G(\overline{p})-G(p)}{G(\overline{p})-G(\hat{p})}\right) & if\ \tilde{p}_{VH}(e)<p<\tilde{p}_H(e)\ \ and\ e\geq\underline{e} \\ f_H(\theta_H)\left(\frac{G(\overline{p})-G(p)}{G(\overline{p})-G(\hat{p})}\right)+ & if\ p<\tilde{p}_{VH}(e)\ \ and\ e\geq\underline{e} \\ +\left(1-f_H(\theta_H)\right)\cdot & \\ \cdot f_V(\theta_V)\left(\frac{G(\hat{p})-G(p)}{G(\hat{p})-G(\underline{p})}\right) & \\ 0 & otherwise, \end{cases}$$

where $\omega=(\theta_V,\theta_H,G)$ is a vector of aggregate labor market variables.

Wages

The FOC of the wage bargaining problem is:

$$\gamma\left(\hat{W}(e,p,w)-U(e)\right)^{\gamma-1}\left(\hat{J}(e,p,w)-V(p)\right)^{1-\gamma}\frac{\partial\hat{W}(e,p,w)}{\partial w}= \\ -(1-\gamma)\left(\hat{W}(e,p,w)-U(e)\right)^{\gamma}\left(\hat{J}(e,p,w)-V(p)\right)^{-\gamma}\frac{\partial\hat{J}(e,p,w)}{\partial w},$$

or simply

$$\gamma\left(\hat{J}(e,p,w)-V(p)\right)\frac{\partial\hat{W}(e,p,w)}{\partial w}=-(1-\gamma)\left(\hat{W}(e,p,w)-U(e)\right)\frac{\partial\hat{J}(e,p,w)}{\partial w}.$$

From the value functions we can find that:

$$\frac{\partial\hat{W}(e,p,w)}{\partial w}=-\frac{\partial\hat{J}(e,p,w)}{\partial w}=1,$$

so the equilibrium wage for every match must satisfy the standard sharing rule:

$$\gamma\left(\hat{J}(e,p,w)-V(p)\right)=(1-\gamma)\left(\hat{W}(e,p,w)-U(e)\right).$$

Balances

The inflow from unemployment can be written as:

$$i_U(e,p)=\begin{cases} f_V(\theta_V)\frac{g(p)}{v_V}u(e) & \text{if } e<\hat{e},\ p<\hat{p} \\ f_H(\theta_H)\frac{g(p)}{v_H}u(e) & \text{if } e\geq\hat{e},\ p\geq\hat{p} \\ \left(1-f_H(\theta_H)\right)f_V(\theta_V)\frac{g(p)}{v_V}u(e) & \text{if } e\geq\hat{e},\ p<\hat{p} \\ 0 & \text{otherwise,} \end{cases}$$

where the first condition is satisfied when both workers and firms use only the vacancy channel; the second condition is satisfied when both workers and firms use the headhunter channel; and the third condition is satisfied when a worker searches through both channels and a firm hires through the vacancy channel.

Similarly, the inflow from employment can be written as:

$$i_E(e,p) = \begin{cases} f_V(\theta_V)\frac{g(p)}{v_V}\int_{\underline{p}}^{\min\{p,\tilde{p}_V(e)\}}\phi(e,p')\,dp' & \text{if } e<\hat{e},\ p<\hat{p} \\ f_H(\theta_H)\frac{g(p)}{v_H}\int_{\underline{p}}^{\min\{p,\tilde{p}_H(e)\}}\phi(e,p')\,dp' & \text{if } e\geq\hat{e},\ p\geq\hat{p} \\ (1-f_H(\theta_H))\,f_V(\theta_V)\cdot & \text{if } e\geq\hat{e},\ p<\hat{p} \\ \cdot\frac{g(p)}{v_V}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')\,dp' & \\ 0 & \text{otherwise.} \end{cases}$$

Aggregates

The aggregates that enter the matching functions are determined as follows. The number of unemployed workers searching through the vacancy channel is, simply, the number of unemployed workers:

$$u_V = \int_{\underline{e}}^{\overline{e}} 1dU(e).$$

The number of unemployed workers searching through the headhunter channel is:

$$u_H = \int_{\hat{e}}^{\overline{e}} 1dU(e).$$

The number of employed workers searching through the vacancy channel is:

$$a_V = \int_{\underline{e}}^{\hat{e}}\int_{\underline{p}}^{\overline{p}} 1d\Lambda_V(e,p) + \int_{\hat{e}}^{\overline{e}}\int_{\underline{p}}^{\overline{p}} 1d\Lambda_{VH}(e,p).$$

The number of employed workers searching through the headhunter channel is:

$$a_H = \int_{\hat{e}}^{\overline{e}}\int_{\underline{p}}^{\overline{p}} 1d\Lambda_{VH}(e,p) + \int_{\hat{e}}^{\overline{e}}\int_{\underline{p}}^{\overline{p}} 1d\Lambda_H(e,p).$$

The number of firms using the vacancy channel is:

$$v_V = \int_{\underline{p}}^{\hat{p}} 1dG(p).$$

And the number of firms using the headhunter channel is:

$$v_H = \int_{\hat{p}}^{\overline{p}} 1dG(p).$$

A.2 Conditions for Separating Equilibrium

A.2.1 Firms

For separation equilibrium to exist, it must be true that low-productive firms strictly prefer to hire through the vacancy channel while high-productive firms strictly prefer to hire through the headhunter channel. Consider the lowest-productive firm. The lowest-productive firm will successfully hire a worker only if it is matched with an unemployed worker. It prefers to hire through the vacancy channel if $V_V(\underline{p}) > V_H(\underline{p})$. We can rewrite the values of posting a vacancy in each channel for such firm as

$$
\begin{aligned}
V_V\left(\underline{p}\right) = \quad & -c_{fV}\left(\underline{p}\right) + \beta\Bigg(V\left(\underline{p}\right) + q_V\left(\theta_V\right)\tfrac{u_V}{u_V+a_V}\Bigg(\textstyle\int_{\underline{e}}^{\hat{e}}\left(J\left(\underline{p},e\right) - V\left(\underline{p}\right)\right)dU\left(e\right) \\
& + \left(1 - f_H\left(\theta_H\right)\right)\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\underline{p},e\right) - V\left(\underline{p}\right)\right)dU\left(e\right)\Bigg)\Bigg)
\end{aligned}
$$

and

$$
V_H\left(\underline{p}\right) = \quad -c_{fH}\left(\underline{p}\right) + \beta\Bigg(V\left(\underline{p}\right) + q_H\left(\theta_H\right)\tfrac{u_H}{u_H+a_H}\Bigg(\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\underline{p},e\right) - V\left(\underline{p}\right)\right)dU\left(e\right)\Bigg)\Bigg) \ .
$$

Condition $V_V\left(\underline{p}\right) > V_H\left(\underline{p}\right)$ holds if

$$
\begin{aligned}
& \beta\Bigg(q_V\left(\theta_V\right)\tfrac{u_V}{u_V+a_V}\Bigg(\textstyle\int_{\underline{e}}^{\hat{e}}\left(J\left(\underline{p},e\right) - V\left(\underline{p}\right)\right)dU\left(e\right) \\
& + \left(1 - f_H\left(\theta_H\right)\right)\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\underline{p},e\right) - V\left(\underline{p}\right)\right)dU\left(e\right)\Bigg)\Bigg) - \qquad > \quad -\left(c_{fH}\left(\underline{p}\right) - c_{fV}\left(\underline{p}\right)\right) \ . \\
& -\beta\Bigg(q_H\left(\theta_H\right)\tfrac{u_H}{u_H+a_H}\Bigg(\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\underline{p},e\right) - V\left(\underline{p}\right)\right)dU\left(e\right)\Bigg)\Bigg)
\end{aligned}
$$

This condition will be satisfied if at least one of the following holds: 1) the cost of hiring through the headhunter channel is sufficiently higher than the cost of hiring through the vacancy channel - $c_{fH}\left(\underline{p}\right) > c_{fV}\left(\underline{p}\right)$; 2) the matching rate with unemployed workers is sufficiently higher in the vacancy channel - $q_V\left(\theta_V\right)\frac{u_V}{u_V+a_V} > q_H\left(\theta_H\right)\frac{u_H}{u_H+a_H}$; 3) there are relatively few unemployed workers above the headhunter threshold; 4) the production of the match increases relatively slowly with the worker's skill for this firm.

Consider now the highest-productive firm. The highest-productive firm must strictly prefer to hire through the headhunter channel, $V_H\left(\overline{p}\right) > V_H\left(\overline{p}\right)$. We can rewrite the vacancy values of the highest-productive firm using the fact that any worker matched with this firm will accept the match.

$$
\begin{aligned}
V_V\left(\overline{p}\right) = \quad & -c_{fV}\left(\overline{p}\right) + \beta\Bigg(V\left(\overline{p}\right) + q_V\left(\theta_V\right)\Bigg(\tfrac{u_V}{u_V+a_V}\textstyle\int_{\underline{e}}^{\hat{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dU\left(e\right) \\
& + \tfrac{u_V}{u_V+a_V}\left(1 - f_H\left(\theta_H\right)\right)\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dU\left(e\right) \\
& + \tfrac{a_V}{u_V+a_V}\textstyle\int_{\underline{e}}^{\hat{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dL_V\left(e\right) \\
& + \tfrac{a_V}{u_V+a_V}\left(1 - f_H\left(\theta_H\right)\right)\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dL_{VH}\left(e\right)\Bigg)\Bigg)
\end{aligned}
$$

and

$$
\begin{aligned}
V_H\left(\overline{p}\right) = \quad & -c_{fH}\left(\overline{p}\right) + \beta\Bigg(V\left(\overline{p}\right) + q_H\left(\theta_H\right)\Bigg(\tfrac{u_H}{u_H+a_H}\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dU\left(e\right) \\
& + \tfrac{a_H}{u_H+a_H}\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dL_H\left(e\right) \\
& + \tfrac{a_H}{u_H+a_H}\textstyle\int_{\hat{e}}^{\overline{e}}\left(J\left(\overline{p},e\right) - V\left(\overline{p}\right)\right)dL_{VH}\left(e\right)\Bigg)\Bigg).
\end{aligned}
$$

Condition $V_H(\overline{p}) > V_H(\overline{p})$ holds if

$$\begin{aligned}
&\beta\Bigg(q_H(\theta_H)\Bigg(\tfrac{u_H}{u_H+a_H}\int_{\hat{e}}^{\overline{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dU(e) \\
&+\tfrac{a_H}{u_H+a_H}\int_{\hat{e}}^{\overline{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dL_H(e) \\
&+\tfrac{a_H}{u_H+a_H}\int_{\hat{e}}^{\overline{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dL_{VH}(e)\Bigg)\Bigg)- \\
&-\beta\Bigg(q_V(\theta_V)\Bigg(\tfrac{u_V}{u_V+a_V}\int_{\underline{e}}^{\hat{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dU(e) \qquad > \quad -\left(c_{fV}(\overline{p})-c_{fH}(\overline{p})\right). \\
&+\tfrac{u_V}{u_V+a_V}\left(1-f_H(\theta_H)\right)\int_{\hat{e}}^{\overline{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dU(e) \\
&+\tfrac{a_V}{u_V+a_V}\int_{\underline{e}}^{\hat{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dL_V(e) \\
&+\tfrac{a_V}{u_V+a_V}\left(1-f_H(\theta_H)\right)\int_{\hat{e}}^{\overline{e}}\left(J(\overline{p},e)-V(\overline{p})\right)dL_{VH}(e)\Bigg)\Bigg)
\end{aligned}$$

This condition will be satisfied, instead, if at least one of the following holds: i) the cost of hiring through the headhunter channel is not much higher than the cost of hiring through the vacancy channel; ii) the absolute matching rate is not much higher in the vacancy channel; iii) there are enough workers accepting the headhunter's calls; iv) the production of the match increases sufficiently fast with the worker's skill for this firm.

A.2.2 Workers

On the worker side, for the separating equilibrium to exist, the highest-skilled worker employed in the lowest-productive firm must agree to consider an offer by a headhunter. Formally, the value of search only through vacancy channel cannot be optimal for such worker: $S_{EVH}\left(\overline{e},\underline{p}\right) > S_{EV}\left(\overline{e},\underline{p}\right)$ and/or $S_{EH}\left(\overline{e},\underline{p}\right) > S_{EV}\left(\overline{e},\underline{p}\right)$.

The values of search for such worker are:

$$\begin{aligned}
S_{EVH}\left(\overline{e},\underline{p}\right) = \;& f_H(\theta_H)\left(\int_{\hat{p}}^{\overline{p}}\left(W(\overline{e},p')-W\left(\overline{e},\underline{p}\right)\right)dG(p')-c_{wH}(\overline{e})\right) \\
&+f_V(\theta_V)\left(1-f_H(\theta_H)\right)\int_{\underline{p}}^{\hat{p}}\left(W(\overline{e},p')-W\left(\overline{e},\underline{p}\right)\right)dG(p') \\
&-c_{wV}(\overline{e}),
\end{aligned}$$

$$S_{EV}\left(\overline{e},\underline{p}\right) = f_V(\theta_V)\int_{p}^{\hat{p}}\left(W(\overline{e},p')-W\left(\overline{e},\underline{p}\right)\right)dG(p')-c_{wV}(\overline{e}),$$

$$S_{EH}\left(\overline{e},\underline{p}\right) = f_H(\theta_H)\left(\int_{\hat{p}}^{\overline{p}}\left(W(\overline{e},p')-W\left(\overline{e},\underline{p}\right)\right)dG(p')-c_{wH}(\overline{e})\right).$$

Start with the first case. Condition $S_{EVH}\left(\overline{e},\underline{p}\right) > S_{EV}\left(\overline{e},\underline{p}\right)$ holds if

$$\begin{aligned}
&f_H(\theta_H)\Bigg(\int_{\hat{p}}^{\overline{p}}\left(W(\overline{e},p')-W\left(\overline{e},\underline{p}\right)\right)dG(p')-c_{wH}(\overline{e})- \\
&-f_V(\theta_V)\int_{\underline{p}}^{\hat{p}}\left(W(\overline{e},p')-W\left(\overline{e},\underline{p}\right)\right)dG(p')\Bigg) \qquad > \quad 0
\end{aligned}$$

This condition will be satisfied if at least one of the following holds: 1) the cost of the headhunter channel is not too high; 2) the matching rate in the vacancy channel is not too high; 3) there are enough firms hiring through the headhunter channel; 4) the production of the match increases sufficiently fast with the firm's productivity for this worker.

The second condition, $S_{EH}\left(\overline{e},\underline{p}\right) > S_{EV}\left(\overline{e},\underline{p}\right)$, holds if

$$\begin{array}{ll} f_H\left(\theta_H\right)\int_{\hat{p}}^{\overline{p}}\left(W\left(\overline{e},p'\right)-W\left(\overline{e},\underline{p}\right)\right)dG\left(p'\right)- & \\ -f_V\left(\theta_V\right)\int_{\underline{p}}^{\hat{p}}\left(W\left(\overline{e},p'\right)-W\left(\overline{e},\underline{p}\right)\right)dG\left(p'\right)- & > \quad 0 \\ -\left(f_H\left(\theta_H\right)c_{wH}\left(\overline{e}\right)-c_{wV}\left(\overline{e}\right)\right) & \end{array}$$

This condition will be satisfied if at least one of the following holds: i) the cost of the headhunter channel is not too high relative to the vacancy channel; ii) the matching rate in the vacancy channel is not too high relative to the headhunter channel; iii) there are enough firms hiring through the headhunter channel; iv) the production of the match increases sufficiently fast with the firm's productivity for this worker.

There is no condition for low-skilled workers in this case because they are excluded from the headhunter channel by assumption.

A.3 Idiosyncratic Costs of Headhunters

First, it is more convenient to define two measures for firms with an open position - $G_V\left(p\right)$ for the firms using vacancy channel, and $G_H\left(p\right)$ for the firms using the headhunter channel.

Workers

The value functions are the following. For low-skilled unemployed workers:

$$S_U\left(e\right)=S_{UV}\left(e\right)\equiv f_V\left(\theta_V\right)\int_{\underline{p}}^{\overline{p}}\left(W\left(e,p\right)-U\left(e\right)\right)dG_V\left(p\right)-c_{wV}.$$

For high-skilled unemployed workers:

$$\begin{aligned} S_U\left(e\right)=S_{UVH}\left(e\right)\equiv\; & f_H\left(\theta_H\right)\left(1-f_V\left(\theta_V\right)\right)\cdot \\ & \cdot\left(\int_{\hat{p}}^{\overline{p}}\left(W\left(e,p\right)-U\left(e\right)\right)dG_H\left(p\right)-c_{wH}\right)-c_{wV}+ \\ & +f_V\left(\theta_V\right)\left(1-f_H\left(\theta_H\right)\right)\int_{\underline{p}}^{\overline{p}}\left(W\left(e,p\right)-U\left(e\right)\right)dG_V\left(p\right)+ \\ & +f_H\left(\theta_H\right)f_V\left(\theta_V\right)\cdot \\ & \cdot\left(\int_{\hat{p}}^{\overline{p}}\int_{\underline{p}}^{\overline{p}}\left(\max\left\{W\left(e,p\right),W\left(e,p'\right)\right\}-U\left(e\right)\right)dG_H\left(p\right)dG_V\left(p'\right)-c_{wH}\right). \end{aligned}$$

For low skilled employed workers:

$$S_{EV}\left(e,p\right)\equiv f_V\left(\theta_V\right)\int_{p}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_V\left(p'\right)-c_{wV}.$$

For high-skilled employed workers:

$$S_{EH}\left(e,p\right)\equiv f_H\left(\theta_H\right)\left(\int_{\max\{\hat{p};p\}}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_H\left(p'\right)-c_{wH}\right),$$

and:

$$\begin{aligned} S_{EVH}\left(e,p\right)\equiv\; & f_H\left(\theta_H\right)\left(1-f_V\left(\theta_V\right)\right)\cdot \\ & \cdot\left(\int_{\max\{\hat{p};p\}}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_H\left(p'\right)-c_{wV}\right)-c_{wV}+ \\ & +f_V\left(\theta_V\right)\left(1-f_H\left(\theta_H\right)\right)\int_{p}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_V\left(p'\right)+ \\ & +f_H\left(\theta_H\right)f_V\left(\theta_V\right)\cdot \\ & \cdot\left(\int_{\hat{p}}^{\overline{p}}\int_{\underline{p}}^{\overline{p}}\left(\max\left\{\max\left\{W\left(e,p''\right),W\left(e,p'\right)\right\}-W\left(e,p\right);0\right\}\right)dG_H\left(p'\right)dG_V\left(p''\right)-c_{wH}\right). \end{aligned}$$

Firms

The value function of a firm posting a vacancy in this case is:

$$\begin{aligned} V_V(p) = \quad & -c_{fV}\cdot p+\beta\Bigg(V(p)+q_V(\theta_V)\Bigg(\tfrac{u_V}{u_V+a_V}\int_{\underline{e}}^{\hat{e}}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dU(e)+ \\ & +\tfrac{u_V}{u_V+a_V}\left(1-f_H(\theta_H)\right)\int_{\hat{e}}^{\overline{e}}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dU(e)+ \\ & +\tfrac{u_V}{u_V+a_V}f_H(\theta_H)\tfrac{G_H(p)}{G_H(\overline{p})}\int_{\hat{e}}^{\overline{e}}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dU(e)+ \\ & +\tfrac{a_V}{u_V+a_V}\int_{\underline{e}}^{\hat{e}}\tfrac{\Lambda_V(e,p)}{\Lambda_V(e,\overline{p})}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dL_V(e)+ \\ & +\tfrac{a_V}{u_V+a_V}\left(1-f_H(\theta_H)\right)\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dL_{VH}(e)+ \\ & +\tfrac{a_V}{u_V+a_V}f_H(\theta_H)\tfrac{G_H(p)}{G_H(\overline{p})}\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dL_{VH}(e)\Bigg)\Bigg). \end{aligned}$$

The value function of a firm using the headhunter channel is:

$$\begin{aligned} V_H(p) = \quad & -c_{fH}\cdot p+ \\ & +\beta\Bigg(V(p)+q_H(\theta_H)\cdot \\ & \cdot\Bigg(\tfrac{u_H}{u_H+a_H}\left(1-f_V(\theta_V)\right)\int_{\hat{e}}^{\overline{e}}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dU(e)+ \\ & +\tfrac{u_H}{u_H+a_H}f_V(\theta_V)\tfrac{G_V(p)}{G_V(\overline{p})}\int_{\hat{e}}^{\overline{e}}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dU(e)+ \\ & +\tfrac{a_H}{u_H+a_H}\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_H(e,p)}{\Lambda_H(e,\overline{p})}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dL_H(e) \\ & +\tfrac{a_H}{u_H+a_H}\left(1-f_V(\theta_V)\right)\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dL_{VH}(e)+ \\ & +\tfrac{a_H}{u_H+a_H}f_V(\theta_V)\tfrac{G_V(p)}{G_V(\overline{p})}\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}\left(J(p,e)-V\left(p,c'_{fN}\right)\right)dL_{VH}(e)\Bigg)\Bigg). \end{aligned}$$

And the value of an open position is:

$$\tilde{V}(p,c_{fN})=\max\left\{V_V(p)\,;V_H(p)-c_{fN}\right\}.$$

The quit rate is:

$$s_Q(e,p,\omega)=\begin{cases} f_V(\theta_V)\left(\frac{G_V(\overline{p})-G_V(p)}{G_V(\overline{p})}\right) & if\ \ p<\tilde{p}_V(e)\ \ and\ e<\underline{e} \\ f_H(\theta_H)\left(\frac{G_H(\overline{p})-G_H(p)}{G_H(\overline{p})}\right) & if\ \ \tilde{p}_{VH}(e)<p<\tilde{p}_H(e) \\ & and\ e\geq\underline{e} \\ (1-f_V(\theta_V))\cdot & if\ \ p<\tilde{p}_{VH}(e)\ \ and\ e\geq\underline{e} \\ \cdot f_H(\theta_H)\left(\frac{G_H(\overline{p})-G_H(p)}{G_H(\overline{p})}\right)+ & \\ +(1-f_H(\theta_H))\cdot & \\ \cdot f_V(\theta_V)\left(\frac{G_V(\overline{p})-G_V(p)}{G_V(\overline{p})}\right)+ & \\ +f_V(\theta_V)f_H(\theta_H)\cdot & \\ \cdot\left(1-\frac{G_V(p)}{G_V(\overline{p})}\frac{G_H(p)}{G_H(\overline{p})}\right) & \\ 0 & otherwise. \end{cases}$$

Aggregation

The number of firms using the vacancy channel is:

$$v = \int_{\underline{p}}^{\hat{p}} 1 dG_V(p).$$

The number of firms using the headhunter channel is:

$$h = \int_{\hat{p}}^{\overline{p}} 1 dG_H(p).$$

The number of searching workers is determined as before.

Balance

The aggregate balance equation is, as before:

$$\phi(e,p)(s + s_Q(e,p)(1-s)) = i_E(e,p) + i_U(e,p),$$

while the inflows now are:

$$i_U(e,p) = \begin{cases} f_V(\theta_V)\frac{g_V(p)}{v_V}u(e) & \text{if } e < \hat{e} \\ f_H(\theta_H)(1 - f_V(\theta_V))\frac{g_H(p)}{v_H}u(e) + & \text{if } e \geq \hat{e} \\ +(1 - f_H(\theta_H)) f_V(\theta_V)\frac{g_V(p)}{v_V}u(e) + & \\ +f_H(\theta_H) f_V(\theta_V)\left(\frac{g_V(p)}{v_V}\frac{G_H(p)}{G_H(\overline{p})} + \frac{g_H(p)}{v_H}\frac{G_V(p)}{G_V(\overline{p})}\right)u(e) & \end{cases}$$

and:

$$i_E(e,p) = \begin{cases} f_V(\theta_V)\frac{g_V(p)}{v_V}\int_{\underline{p}}^{\min\{p,\tilde{p}_V(e)\}}\phi(e,p')\,dp' & \text{if } e < \hat{e} \\ f_H(\theta_H)\frac{g_H(p)}{v_H}\int_{\min\{p,\tilde{p}_{VH}(e)\}}^{\min\{p,\tilde{p}_H(e)\}}\phi(e,p')\,dp' + & \text{if } e \geq \hat{e} \\ +f_H(\theta_H)(1 - f_V(\theta_V))\frac{g_H(p)}{v_H}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')\,dp' + & \\ +(1 - f_H(\theta_H)) f_V(\theta_V)\frac{g_V(p)}{v_V}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')\,dp' + & \\ +f_H(\theta_H) f_V(\theta_V)\frac{g_V(p)}{v_V}\frac{G_H(p)}{G_H(\overline{p})}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')\,dp' + & \\ +f_H(\theta_H) f_V(\theta_V)\frac{g_H(p)}{v_H}\frac{G_V(p)}{G_V(\overline{p})}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')\,dp' & \end{cases}$$

A.4 Robustness of Quantitative Results

To check how much the magnitude of the increase of top wages depends on the choice of the skill threshold for the headhunter channel, I do similar experiments for top 1%, 2.5%, 5%, 7%, or 10% of workers being eligible for the headhunter channel. The results are presented in Table 9. Not surprisingly, a higher threshold increases the wage share of the top 1% of the workers but decreases the share of the top 10%. This happens because, with a higher threshold, the most efficient workers are more concentrated in the top firms, for example, they all work in top 2.5% of the firms instead of top 5%. Their wages increase even more due to complementarities, so the top 1% wage share increases more. Instead, for the workers in the 10-1% bracket, the probability of working in the best firms decreases with a higher threshold. Workers in 5-2.5% are excluded from the headhunter channel and many of them end up in bad or average firms, so the top 10% wage share drops relative to the baseline calibration despite the top 1% wage share increase.

Another target that doesn't have a properly estimated empirical counterpart is the share of firms using the headhunters. I redo the experiment with different shares to see

Table 9: Top Wage Shares in the Model for Different Skill Thresholds

Model	Share	Top 1%	Top 10%	90/50
Without HH	0%	4.97%	27.84%	2.42
With HH on top 17% (basline)	16.9%	5.61%	30.96%	2.74
With HH on top 10%	11.4%	5.82%	30.53%	2.54
With HH on top 7%	7.2%	5.95%	30.34%	2.43
With HH on top 5%	4.2%	6.22%	29.82%	2.42
With HH on top 2.5%	2.3%	7.30%	29.68%	2.46
With HH on top 1%	1.2%	6.57%	28.59%	2.49

how sensitive are the results depending on the choice of the target. I set the share of the firms using the headhunter channel to be 10%, 20%, 40%, 60%, 80%, or 100%. The results are presented in Table 10. The increase of the top 10% wage share is decreasing with a lower share of firms using headhunters, but the major part of the effect is still there even if every 10th firm is allowed to use the headhunter channel every period. In this case, the model is still able to explain 36% of the increase in the top 10% wage share (together with SBTC). Even when the share of firms using headhunters is set to the most conservative estimate, the model is still able to predict a large share of the increase in top wages.

Table 10: Top Wage Shares in the Model for Different Headhunter Intensity

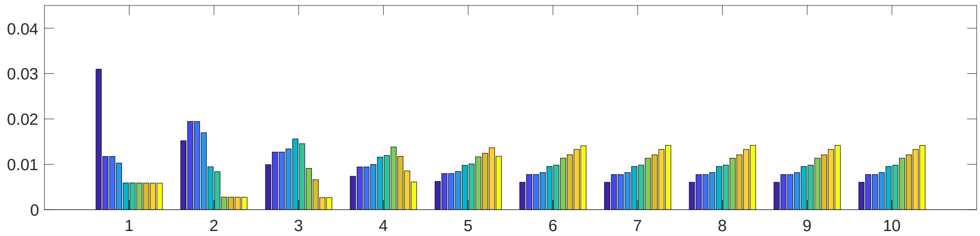
Model	Top 1%	Top 10%	90/50
Without HH	4.97%	27.84%	2.42
With HH, baseline	5.61%	30.96%	2.74
With HH, 100%	6.45%	37.65%	3.50
With HH, 80%	6.17%	33.46%	3.21
With HH, 60%	5.96%	31.93%	2.80
With HH, 40%	5.51%	31.34%	2.74
With HH, 20%	5.41%	30.34%	2.74
With HH, 10%	5.14%	29.05%	2.63

A.5 Additional Figures

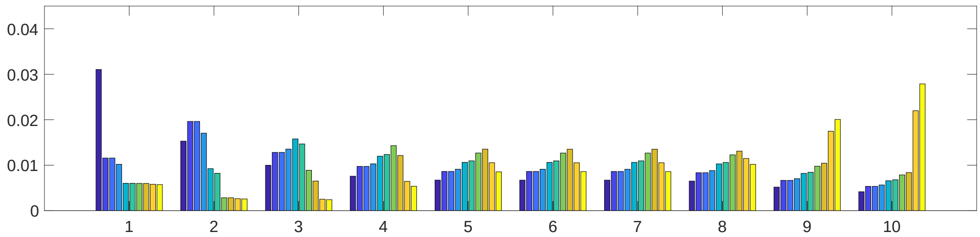
Figure 9a shows the joint distribution of workers and firms in the model without the headhunter channel, Figure 9b shows the distribution with the headhunter channel, and Figure 9c shows the change of the distribution. Numbers 1,2,3,..,10 in the figure correspond to the firm types, with 1 being the least productive firms and 10 being the most productive firms, and the colors correspond to the types of workers, with dark blue being the least skilled and yellow being the most skilled workers.

A.6 Additional Evidence on Headhunter Industry

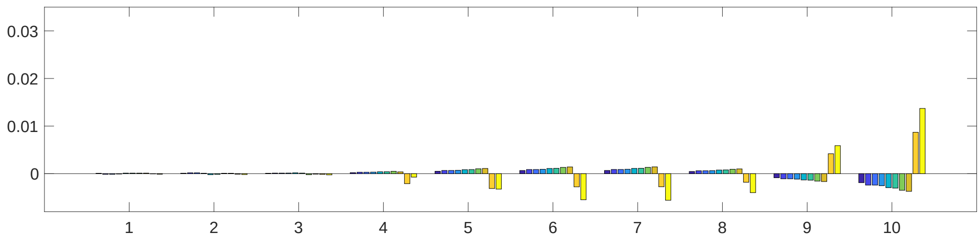
How are the revenues of the headhunter industry distributed? Headhunters cover a wide range of positions: CEOs, board directors, CFOs, senior executives, general management, top professionals in finance and control, information systems, marketing, and sales. They



(a) Joint Distribution without the Headhunter Channel



(b) Joint Distribution with the Headhunter Channel



(c) Change in the Joint Distribution

Figure 9: Joint Distributions of Worker-Firm Matches

are not focused only on CEOs, as is sometimes perceived, but cover almost all the top positions in companies. The industry composition of headhunter operations is also dispersed; they operate in all industries as illustrated by the distribution of fee revenues. Distribution of fee revenues by industry in the 4th quarter of 2015 is presented in Table 11.

Table 11: Fee Revenues of Headhunters by Industry, 4th Quarter 2015, from AESC

Industry	Share (%)
Industrial	23.5
Financial	21.0
Consumer products	18.0
Technology	16.2
Life Sciences/Healthcare	15.2
Non-Profit	4.4
Other	1.7

Source: AESC Insights Q4 2015 State of the Executive Search Industry.

Geographical distribution of fee revenues, instead, is not so homogeneous, as can be seen in Table 12. Headhunters receive most of the revenue from North America, and mainly the U.S. Europe is lagging behind and the major part of European revenues comes from the U.K. There might be several reasons for such difference. One possible reason is the difference in labor market legislation. It is more difficult to be an intermediary in a European labor market than in the U.S. Another possible reason is the cost of creating a database of potential candidates in a new country. The headhunter must know the specifics of the labor market and the companies operating in the country in order to understand the skills demanded by companies as well as the value of observable signals, such as particular diplomas and experience in particular companies. Because headhunters first appeared in the U.S. they established the databases and acquired the knowledge of the labor market and the signals there first.

Table 12: Fee Revenues of Headhunters by Region, 4th Quarter 2015, from AESC

Region	Share (%)
North America	45.5
EMEA	33.2
Asia Pacific	16.5
Latin America	4.8

Source: AESC Insights Q4 2015 State of the Executive Search Industry.

A.7 COMPUSTAT Data

The four firm size proxies are constructed as follows. First, firm value is constructed as the sum of the market value of equity, defined as a number of shares outstanding multiplied by the end-of-fiscal-year stock price, and the book value of debt, defined as total assets minus the sum of the book value of equity and deferred taxes. Second, equity value constructed as the number of shares outstanding multiplied by the end-of-fiscal-year stock price. Third, the sales variable from the COMPUSTAT. Fourth, the income is measured as earnings before interest and taxes.

A.8 Additional Micro Evidence

Table 13 presents the results for individual measures of the firm size. Columns (1) and (2) present the results for the firm value measure and columns (3) and (4) for the equity value measures. The results are consistent with the results for the full sample.

Table 13: CEO Compensation and the Change of the CEO, Individual Firm Size Measures

	Sample period 1993-2013			
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.0437	0.1389	0.0420	0.1876
	(0.0177)	(0.0297)	(0.0171)	(0.0314)
Log of Firm Value	0.4311	0.4620	-	-
	(0.0180)	(0.0193)	-	-
Log of Equity Value	-	-	0.3442	0.3572
	-	-	(0.0161)	(0.0191)
Year FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
R^2	0.66	0.653	0.66	0.654
Number of observations	24673	24673	24673	24673

Table 14 presents the results of the estimation using an interaction term between measures of firm size and the new CEO dummy (as a proxy for the bargaining power). Columns (1) and (2) present the results for the firm value as a proxy for the firm size and columns (3) and (4) present the results for the equity value of the firm.

Table 14: CEO Compensation and the Change of the CEO, Bargaining Power

	Sample period 1993 - 2013			
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.3590	0.3145	0.4079	0.3168
	(0.1293)	(0.1277)	(0.1318)	(0.1302)
Log of Firm Value	0.4746	0.4466	-	-
	(0.0173)	(0.0183)	-	-
Log of Equity Value	-	-	0.3720	0.3629
	-	-	(0.0159)	(0.0164)
Log of FV*New CEO	-0.0269	-0.0332	-	-
	(0.0165)	(0.0165)	-	-
Log of EV*New CEO	-	-	-0.0291	-0.0366
	-	-	(0.0182)	(0.0184)
Year FE	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes
R^2	0.654	0.659	0.654	0.662
Number of observations	24673	24673	24673	24673

To evaluate the matching channel, I test the effect of a change of the CEO on the firm size. Table 15 presents the results, columns (1) and (2) show the effect on the firm value and columns (3) and (4) show the effect on the equity value. As it can be seen from the table, the effect of the change of the CEO on the firm size is positive. These results possibly indicate the presence of the channel related to the productivity of the match between the new CEO and the firm.

Table 15: CEO Compensation and the Change of the CEO, Match Efficiency

	Firm Value		Equity Value	
	(1)	(2)	(3)	(4)
New CEO	0.4784	0.6138	0.4023	0.5208
	(0.0399)	(0.0450)	(0.0399)	(0.0448)
Year FE	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes
R^2	0.200	0.231	0.095	0.132
Number of observations	24673	24673	24673	24673

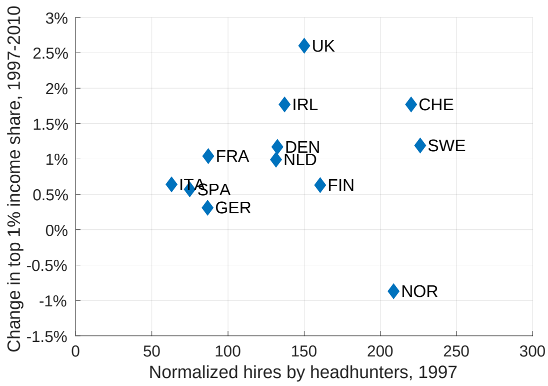
Table 16 presents the results of the regression including just the enforceability index, but not the new CEO dummy. the results confirm the well-known result that CEO compensation is lower in the states that enforce the non-compete agreements.

Table 16: CEO Compensation and the Non-Compete Enforceability Index

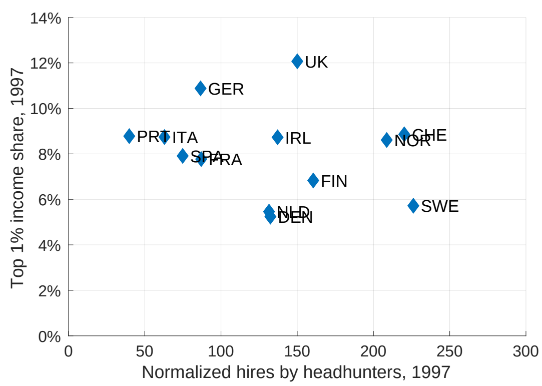
	Sample period 1993 - 2013		
Log of compensation	(1)	(2)	(3)
NCEI	-0.0222	-0.0286	-0.0221
	(0.0028)	(0.0029)	(0.0029)
Firm Size Controls	Yes	Yes	Yes
Year FE	No	Yes	Yes
Industry FE	Yes	No	Yes
R^2	0.410	0.384	0.420
Number of observations	24217	24217	24217

A.9 Additional Cross-Country Evidence

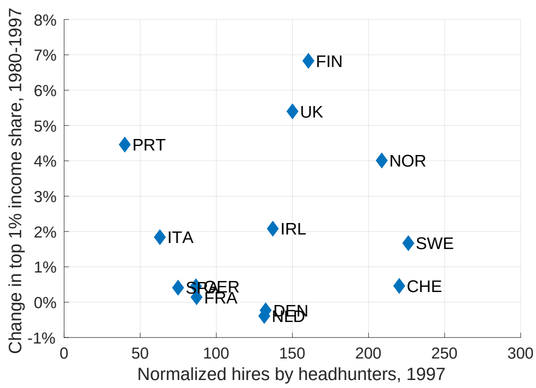
Figure 10 repeats the analysis presented in Section 5 for the case of the top 1% income shares. Figures 11 and 12 plot similar relations for normalized fee revenues. It is evident from the figures that the pattern stays the same regardless of the measure used in the analysis. Both measures of headhunter intensity predict future growth in both the top 1% and the top 10% income shares while there is no positive correlation between these measures and the level of top incomes or previous top income growth.



(a) Top 1% Income Share Growth, 1997-2010, and Normalized Hires by Headhunters, 1997

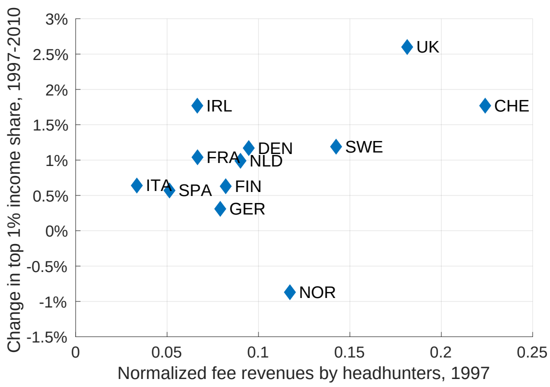


(b) Top 1% Income Share, 1997, and Normalized Hires by Headhunters, 1997

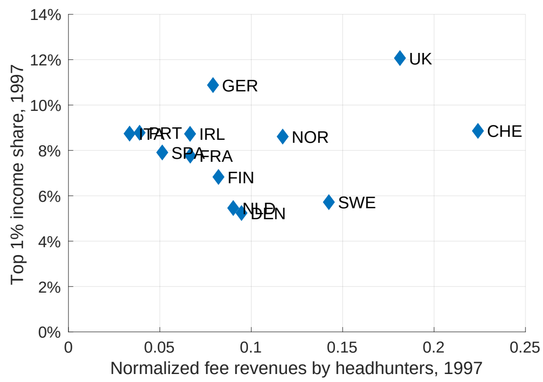


(c) Top 1% Income Share Growth, 1980-1997, and Normalized Hires by Headhunters, 1997.

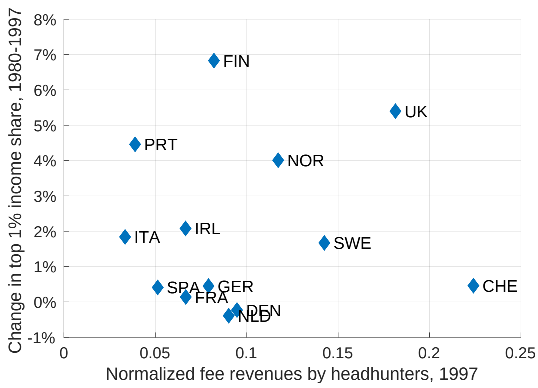
Figure 10: Top 1% Income Share and Normalized Hires by Headhunters



(a) Top 1% Income Share Growth, 1997-2010, and Normalized Fee Revenues by Headhunters, 1997

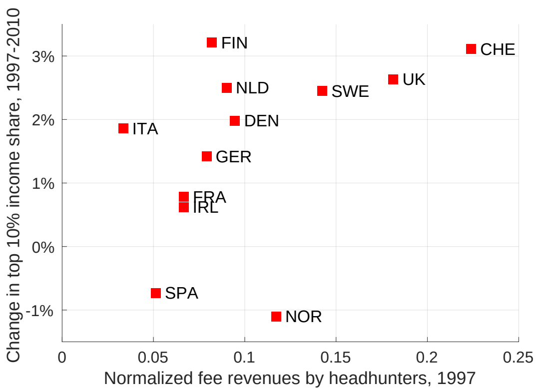


(b) Top 1% Income Share, 1997, and Normalized Fee Revenues by Headhunters, 1997

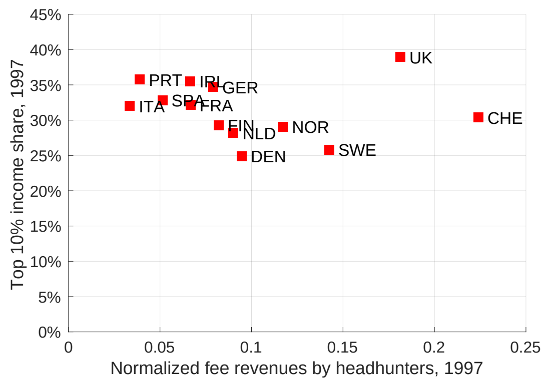


(c) Top 1% Income Share Growth, 1980-1997, and Normalized Fee Revenues by Headhunters, 1997

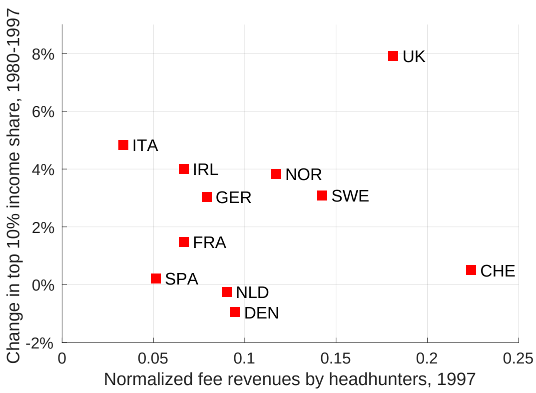
Figure 11: Top 1% Income Share and Normalized Fee Revenues by Headhunters



(a) Top 10% Income Share Growth, 1997-2010, and Normalized Fee Revenues by Headhunters, 1997



(b) Top 10% Income Share, 1997, and Normalized Fee Revenues by Headhunters, 1997



(c) Top 10% Income Share Growth, 1980-1997, and Normalized Fee Revenues by Headhunters, 1997

Figure 12: Top 10% Income Share and Normalized Fee Revenues by Headhunters