An Empirical Model of Wage Dispersion with Sorting*

Jesper Bagger[†] Rasmus Lentz[‡]

April 26, 2017

Abstract

We estimate an equilibrium on-the-job search model with endogenous search intensity. Workers differ by skill, firms by productivity. Workers respond to mismatch by intensive search, and sorting may result from complementarities in the match-level production function. The model is estimated on Danish matched employer-employee data. Firms are ranked through revealed preference by the fraction of hires that is poached from other firms: The poaching rank. Identification is obtained by firm rank conditional mobility and wage patterns. Wage variation is decomposed into four sources: Sorting (40%), worker heterogeneity (30%), firm heterogeneity (20%), and frictional competition (10%). A social planner can improve output net of search cost by 1.4% relative to the decentralized solution.

Keywords: Sorting, Search Intensity, Worker heterogeneity, Firm heterogeneity, On-the-job search, Mismatch, Wage dispersion, Matched employer-employee data

JEL codes: J24, J33, J62, J63, J64, D83, E24

*We have benefited from comments and suggestions from Jan Eeckhout, John Kennan, Philipp Kircher, Thibaud Lamadon, Jeremy Lise, Costas Meghir, Rafael Lopes de Melo, Dale T. Mortensen, Giuseppe Moscarini, Fabien Postel-Vinay, Jean-Marc Robin, Richard Rogerson, Robert Shimer, Chris Taber as well as three anonymous referees and our editor. Henning Bunzel has been invaluable in the development of the data we use. We acknowledge financial support from LMDG and CAP, both Aarhus University research centers. Rasmus Lentz acknowledges support from NSF grant SES-1459897.

[†]Royal Holloway, University of London; LMDG; CAP; E-mail: jesper.bagger@rhul.ac.uk

[‡]University of Wisconsin-Madison; LMDG; CAP; E-mail: rlentz@wisc.edu

1 Introduction

Why do wages differ across jobs? There is ample empirical evidence that worker skill and firm productivity heterogeneity are both important contributors to observed wage dispersion, see e.g. Abowd et al. (1999). As argued in Burdett and Judd (1983) and Burdett and Mortensen (1998), labor market friction provides a separate source of wage dispersion and allows firm heterogeneity to manifest itself in wages. Postel-Vinay and Robin (2002) and Cahuc et al. (2006) show that these effects are of first order importance.

The matched employer-employee data used in studies such as these also reveal that labor markets are characterized by a large amount of worker reallocation, much of which happens through job-to-job transitions. Reallocation of labor between firms is a costly but also productivity enhancing activity which in a decentralized economy is guided by wages. Indeed, job-to-job transitions tend to be in the direction of higher wages and more productive matches.[1] Therefore, the study of wage dispersion should contain an understanding of the allocation of workers to firms and how the allocation is obtained. In particular, this includes the issue of labor market sorting.

We quantify the sources of wage dispersion in an equilibrium on-the-job search model with firm and worker heterogeneity and the possibility of sorting as workers respond to skill dependent variation in incentives to reallocate to more productive matches. Heterogeneity is single dimensional. Workers differ by skill, firms by productivity. Sorting is driven by complementarities in the match level production technology that combines worker skill and firm productivity. Wages are set as in Dey and Flinn (2005) and Cahuc et al. (2006). A worker's current wage depends on her skill level, her employer's productivity as well as her bargaining position. The latter is shorthand for the competition that arises between two firms when a currently employed worker meets another firm. Meetings are chance events, and individual bargaining positions, and therefore wages, evolve stochastically, even among identical workers employed in identical firms. The model delivers a structural log wage variance decomposition with four components: Worker and firm heterogeneity, frictional competition (defined as within-match bargaining position variation), and sorting.

Using Danish matched employer-employee data we find evidence of positive sorting in the labor market. The estimated equilibrium allocation of workers to firms implies a correlation between worker skill and firm productivity of 0.37. Sorting is the single biggest contributor to wage dispersion, with 40% of the variance in log wages attributed to the positive association between worker

[1] See for example Fallick and Fleischman (2001), Christensen et al. (2005), Nagypál (2005) and Jolivet et al. (2006).

skill and firm productivity. By themselves, worker and firm heterogeneity contribute 30% and 20%, respectively. Finally, frictional competition for workers' services in the labor market contributes 10%. We find that the decentralized economy is slightly over-sorted relative to that of the social planner who can improve output net of search cost by 1.4%.[2]

Three quarters of the contribution from sorting is due to a positive within-firm covariance between worker skill and the worker's bargaining position. That is, more skilled workers extract a greater share of rents from given matches. With this, the analysis provides a reconciliation between our results and those in Abowd et al. (1999) and Postel-Vinay and Robin (2002). The former finds that wage dispersion is predominantly explained through worker wage fixed effect variation, whereas the latter finds a much smaller contribution from worker heterogeneity and a large contribution from frictional wage dispersion. Our estimate implies roughly the same amount of wage variance from bargaining position variation as in Postel-Vinay and Robin (2002). As such, we agree that a large portion of wage variance is from this channel. However, allowing for sorting in the analysis reveals that a significant part of it is correlated with worker skill which we classify as a sorting contribution. Abowd et al. (1999) has no explicit notion of bargaining position variation. If we attribute the positive within-firm covariance between worker skill and the worker's bargaining position to the worker variance contribution, we obtain a variance decomposition similar to that of Abowd et al. (1999). With this interpretation, a more skilled worker has a high wage fixed effect not just because she is more productive but also because she extracts a greater share of rents.

The correlation between worker and firm wage fixed effects as obtained in an Abowd et al. (1999) log wage regression is essentially zero in the Danish data. It is notable that despite the true correlation of 0.37 between worker skill and firm productivity in our estimated model, it reproduces a worker and firm wage fixed effects correlation that is close to the zero correlation in the data. We understand this as a discrepancy between the true wage process in the model and that implied by the Abowd et al. (1999) specification. For example, the wage function in the estimated model is not monotone in the worker and firm types.[3] As demonstrated in Appendix C, and as also emphasized in Eeckhout and Kircher (2011), Lise et al. (2016), and Lopes de Melo (2016), the correlation between the Abowd et al. (1999) wage fixed effects may in this case provide a significantly downward biased

[2]In the interest of comparison with studies such as Gautier and Teulings (2012) that study the output gain in the absence of frictions for fixed job and worker populations, our estimated economy obtains an 8.1% output improvement for a perfectly sorted allocation of the model's estimated worker and job populations.

[3]The production function is assumed monotonically increasing in each index. The non-monotonicity is a result of sorting in a frictional environment.

estimate of the true match distribution correlation between skill and productivity.

As in the partnership models of Becker (1973) and Shimer and Smith (2000), our analysis links sorting in the match distribution to match production function complementarities, but sorting obtains through a different mechanism. In our setup, due to an assumption of constant returns to scale in production, as well as workers' opportunity to search on the job, neither firms nor workers view the match formation decision to include a search opportunity loss. In contrast, the acceptance/rejection decision that drives assortative matching in Shimer and Smith (2000) relies on a fundamental scarcity: Once matched, the agent gives up the opportunity to search until once again unmatched. This is not an altogether unreasonable assumption in the study of marriage, as in Becker (1973), but the vital role of on the job search and the fact that a single firm can match with many workers as well as possibly engage in replacement hiring make the scarcity assumption in labor market matching less obvious.[4]

We follow Burdett (1978) and let workers choose how intensely to search for outside job opportunities. Christensen et al. (2005) show that match distributions in the Danish matched employer-employee data imply that workers reduce search intensity as they move up the firm ladder. Mueller (2010) finds direct evidence in support of this conclusion based on the American Time Use Survey where workers are observed to reduce search intensity given higher wages. Given complementarities in the production function between worker skill and firm productivity, the workers' search intensity choices vary with skill. Positive complementarities in production induce more skilled workers to search more intensely to move up the firm hierarchy. In this case more skilled workers will stochastically be matched with more productive firms, and positive assortative matching obtains. Negative complementarities in production result in negative assortative matching.[5]

Identification of sorting, that is, recovering from observed data the relationship between (possibly unobserved) worker skill and firm productivity, is inherently difficult and is a central question in the paper.[6] The identification strategy in the paper is focused on the mobility part of the data. Firm type rank is identified by revealed preference based on the fraction of a firm's hires that comes

[4]In flexible frameworks that can in principle reveal worker type dependent firm ladders, recent work by Bonhomme et al. (2016), Haltiwanger et al. (2016), and Lentz et al. (2016) have yet to find evidence of disagreement across worker types about firm ranks. It is a key feature of sorting in the partnership model that there is disagreement across worker types about the ranking of firms.

[5] The core sorting mechanism in the model is analyzed in detail in Lentz (2010).

[6]In this respect, it is joined by Gautier and Teulings (2006), Eeckhout and Kircher (2011), Lise et al. (2016), Lopes de Melo (2016), and more recently by Bartolucci and Devicienti (2015) and Hagedorn et al. (2016), all of which argue identification by the adoption of the partnership model assumptions of match opportunity scarcity. These identification strategies are not valid in our setting.

directly from other firms as opposed to from unemployment: The poaching index. With this, identification is obtained through firm rank conditional mobility and wage patterns. We treat worker skill as unobserved but we demonstrate in Section 4 that the data are characterized by positive sorting between firm rank and worker education length. In addition, we also find that better educated workers have shorter unemployment durations pointing to the central mechanism in the paper, that more skilled workers search away from poor matches faster.

The paper is structured as follows: The model is presented in section 2. Sections 3, 4 and 5 present data, key identification arguments and model estimation and fit. Section 6 discusses the implied estimate for efficiency loss due to mismatch and in section 7 we decompose the estimated wage variance into its four distinct sources; worker heterogeneity, firm heterogeneity, imperfect labor market competition, and sorting, and discusses the decomposition in relation to the Abowd et al. (1999) and Postel-Vinay and Robin (2002) contributions. Section 8 concludes.

2 Model

2.1 Production and search technologies

There is a continuum of firms with measure m, and a continuum of risk neutral workers with measure normalized at unity. Time is continuous and firms and workers discount time at a common rate r. Workers maximize income and firms maximize profits. A worker is characterized by his or her permanent innate skill level $h \in [0, 1]$ which is independently and identically distributed across workers according to the cumulative distribution function $\Psi(\cdot)$. Firms differ with respect to their permanent productivity $p \in [0, 1]$ which is independently and identically distributed across firms according to the cumulative distribution function $\Phi(\cdot)$.

Workers can be either employed or unemployed. Regardless of employment state, a worker generates job offers through a choice of search intensity s at increasing and convex cost $c(s)$. A job offer is a draw of a firm productivity from the vacancy distribution $\Gamma(\cdot)$ with pdf $\gamma(\cdot)$. The efficiency of the search technology may differ between employment states. A search intensity s results in an arrival rate of new job opportunities of $(\mu + \kappa s)\lambda(\theta)$ if unemployed and $s\lambda(\theta)$ if employed, where $\kappa > 0$. If $\kappa > 1$ then search is more efficient in the unemployed state. $\mu \geq 0$ represents arrival of offers that is unrelated to the search decision of the worker, normalized by $\lambda(\theta)$. $\lambda(\theta)$ is the equilibrium arrival rate of offers per search unit and θ is the market tightness. By assumption $\lambda'(\theta) \geq 0$. We will often suppress θ in the expression for $\lambda(\theta)$.

A match between a worker of skill level h and a firm of productivity p produces output $f(h,p)$. It is assumed that $f(h,p)$ is twice continuously differentiable with $f_p(h,p) \geq 0$ and $f_h(h,p) \geq 0$ for all (h,p). Hence, more skilled workers enjoy an absolute advantage relative to less skilled workers regardless of the firm type p they are matched with. Likewise for the ranking of firms. We do not restrict the modularity of the production technology. That is, the sign of the cross partial derivative $f_{hp}(h,p)$ is unrestricted. Firms operate at constant returns to scale at the match level. Hence, a firm's output is the sum of the match outputs.

Match separation occurs as the result of one of three mutually exclusive events. First, the worker may receive an offer from an outside firm with greater productivity than the current firm. As we shall see, this will induce a quit. Second, the worker receives an advance notice layoff shock at rate δ_0, normalized by $\lambda(\theta)$. Given the advance notice the worker searches for a new job with the now lower value of the current job. Workers in higher value jobs have a higher reservation threshold and their new job is in expectation a result of more draws from the offer distribution. We adopt a modeling simplification whereby an advance notice layoff shock simply results in a move to another job that is the best offer from $n(p)$ draws from the offer distribution, where $n(p)$ is assumed weakly increasing in p.[7] Furthermore, let the number of draws at the bottom of the firm hierarchy be $n(0) = 1$, reflecting that any new job dominates this job conditional on the worker holding the job in the first place. The special case where $n(p) = 1$ for all $p \in [0,1]$ corresponds to the exogenous reallocation shock in Jolivet et al. (2006).[8] Third, at rate $\delta(p)$ the worker is immediately laid off and moves into unemployment. The layoff rate is allowed to depend on the firm type in such a way that more valuable jobs are less prone to destruction. That is, $\delta(p)$ is assumed weakly decreasing in p. In the remainder of the paper, it is assumed that $\mu = \delta_0$.

2.2 Bargaining and employment contracts.

An employment contract consists of a wage and a search intensity, (w,s), which are fixed until both parties agree to renegotiate. The analysis assumes search intensities can be contracted upon, which results in the implementation of the jointly (between employer and employee) efficient search in-

[7] A related interpretation may be that jobs are subject to idiosyncratic productivity shocks where the distribution of the new productivity draw is stochastically increasing in the current productivity of the match. Either way, a full modeling of the process involves taking a stand on the wage process during the advance notice period or its dependence on idiosyncratic productivity shocks.

[8] Also referred to as the "Godfather shock", as in an offer that cannot be refused.

tensity level, a useful benchmark.[9] The outcome of the employment contract bargaining is such that the agreed upon search intensity maximizes the joint surplus of the match and the wage dictates the surplus split. When unmatched the worker searches so as to maximize the value of unemployment.

Contracts, renegotiated by mutual consent only, are set through a Rubinstein (1982) style bargaining game as in Cahuc et al. (2006). It is assumed that the worker can use a contact with one employer as a threat point in a bargaining game with another. The bargaining procedure is observationally equivalent to the one in Dey and Flinn (2005), which is generalized Nash bargaining with worker bargaining power β. If two firms compete over a worker, the worker moves to, or is retained by, the more productive firm, as the case may be. The worker bargains with full surplus extraction at the less productive firm as the outside option. If the worker is unemployed, then the value of unemployment will be the worker's outside option.

Denote by $V(h,p)$ the joint net present value of match to the worker and the firm where the search intensity $s(h,p)$ is chosen optimally to maximize this object. If a worker is bargaining with a type p firm given an outside option of full surplus extraction from a type $q \leq p$ firm, the outcome gives the worker a net present value of,

$$V(h,q,p) = \beta V(h,p) + (1-\beta)V(h,q), \tag{1}$$

where the wage flow $w(h,q,p)$ is set so as to achieve the value split.

2.3 Employment contract

The joint value to the firm and the worker from a match can be written as,

$$\begin{aligned}\big(r+\delta(p)+\lambda\delta_0\big)V(h,p) = \max_{s\geq 0}\Bigg[& f(h,p) - c(s) + \lambda s\beta\int_p^1 \big[V(h,p')-V(h,p)\big]d\Gamma(p') + \\ & \big(\delta(p)+\lambda\delta_0\big)V_0(h) + \lambda\delta_0\beta\int_{R(h)}^1 \big[V(h,p')-V_0(h)\big]d\Gamma(p')^{n(p)}\Bigg],\end{aligned} \tag{2}$$

where $V_0(h)$ is the worker's valuation of unemployment and $R(h)$ is the worker's reservation threshold defined by $V(h,R(h)) = V_0(h)$. Net of search cost, the match generates output flow $f(h,p) - c(s)$. At rate λs the worker in the match meets an outside firm p', which in case $V(h,p') > V(h,p)$ yields net value of $V(h,p,p') - V(h,p)$, and otherwise zero. At rate $\delta(p)$ the match ends in which

[9] Lentz (2014) studies the mechanism design problem in the case where search intensity is not contractable. Here, a flat wage profile that does not deliver the entire surplus to the worker results in the worker searching too much relative to the jointly efficient level because part of the incentive to generate outside offers now includes rent extraction from the current match.

case the worker moves to unemployment. The advance notice shock also gives the worker a draw from the CDF $\Gamma(\cdot)^{n(p)}$, which is accepted if it exceeds the workers reservation threshold, $R(h)$. The offer acceptance threshold formulation relies on $V(h,p)$ being monotonically increasing in p. This follows from $f(h,p)$ and $n(p)$ increasing in p and the positive slope is amplified by $\delta(p)$ weakly decreasing in p. It is furthermore straightforward to show that there exists a unique solution to the value function. We provide a proof in the appendix. The worker's valuation of unemployment can be written as,

$$rV_0(h) = \max_{s\geq 0}\left[f(h,0) - c(s) + (\mu+\kappa s)\lambda\beta\int_{R(h)}^{1}\big[V(h,p') - V_0(h)\big]d\Gamma(p')\right]. \tag{3}$$

A type-h worker's valuation of a contract with a type p firm, characterized by a wage $w(h,q,p)$ and search intensity $s(h,p)$ can be written recursively by,

$$\begin{aligned}\big(r+\delta(p)+\lambda\delta_0+\lambda s(h,p)\hat{\Gamma}(q)\big)V(h,q,p) = w(h,q,p)-c\big(s(h,p)\big)+\lambda s(h,p)\int_p^1 V(h,p,p')d\Gamma(p')\\ + \lambda s(h,p)\int_q^p V(h,p',p)d\Gamma(p') + \big(\delta(p)+\lambda\delta_0\big)V_0(h)\\ + \lambda\delta_0\beta\int_{R(h)}^1\left[V(h,p') - V_0(h)\right]d\Gamma(p')^{n(p)},\end{aligned} \tag{4}$$

where $\hat{\Gamma}(\cdot) = 1 - \Gamma(\cdot)$ and q represents the worker's bargaining position in the sense of equation (1), that is, full surplus extraction with a type-q firm. If the worker meets a firm of type $p' > p$, the worker leaves and receives a contract with the new firm that yields value $V(h,p,p')$. If the worker meets an outside firm that would be willing to offer greater value than the worker's current contract but cannot offer more than the worker's current firm, the contract is renegotiated and the worker stays. This happens if $q \leq p' \leq p$ in which case the worker receives a new contract with value $V(h,p',p)$. The layoff and advance notice shocks contribute to the worker's contract valuation in ways identical to those in equation (2). Together, equations (1)-(4) determine $w(h,q,p)$.

The optimal search intensity in an (h,p)-match satisfies the first order condition,

$$c'\big(s(h,p)\big) = \lambda\beta\int_p^1 V_p(h,p')d\hat{\Gamma}(p'), \tag{5}$$

where the right hand side represents the expected rents extracted from outside employers from an additional unit of search. Similarly, an unemployed worker chooses search intensity $s_0(h)$ that solves,

$$c'\big(s_0(h)\big) = \kappa\lambda\beta\int_{R(h)}^1 V_p(h,p')d\hat{\Gamma}(p'). \tag{6}$$

It follows directly from the first order condition (5) that $s(h,p)$ is decreasing in p.

In a simpler model where layoff and advance notice shocks do not depend on firm type, Lentz (2010) links variation in search intensity across worker skill to production function complementarities. Specifically, if the production is modular, $f_{hp}(h,p)=0$, then search intensity is invariant in skill for a given firm type, $s_h(h,p)=0$. If the production is globally super (sub) modular then search intensity is increasing (decreasing) in worker skill, that is $s_h(h,p)f_{hp}(h,p)\geq 0$. We numerically verify that our model solution has the same property in our setting with firm type dependent layoff rates and advance notice shocks.

The reservation productivity $R(h)$ is defined by,

$$V(h,R(h))=V_0(h). \tag{7}$$

It is straightforward to show that $R(h)=0$ if $\kappa<1$, which turns out to be the relevant case for the estimated model. The reservation productivity level is non-trivial if unemployed search is more efficient than employed search, $\kappa>1$, where one obtains that $R(h)>0$. However, in the case where the production function is modular, it is straightforward to verify that the reservation threshold is invariant in worker skill.

2.4 Steady state worker flows

Denote by $g(p|h)$ the fraction of employed skill-h workers that are matched with productivity-p firms (by definition, $\int_0^1 g(p|h)dp=1$). Denote by e_h the fraction of skill-h workers that are employed. In steady state, the flow into the stock $e_h g(p|h)$ must equal the flow out, which translates to the condition,

$$\tilde{g}(h,p)=\frac{\gamma(p)\left[1+\int_0^p s(h,p')\tilde{g}(h,p')dp'+\delta_0\int_0^1 n(p')\Gamma(p)^{n(p')-1}\tilde{g}(h,p')dp'\right]}{\delta_0+\hat{\delta}(p)+s(h,p)\widehat{\Gamma}(p)}, \tag{8}$$

where $\hat{\delta}(p)=\delta(p)/\lambda$ and,

$$g(p|h)=\left[\frac{1-e_h}{e_h}\big(\kappa s_0(h)+\mu\big)\right]\tilde{g}(h,p), \tag{9}$$

and

$$e_h=\frac{\big(\kappa s_0(h)+\mu\big)\int_0^1\tilde{g}(h,p)dp}{1+\big(\kappa s_0(h)+\mu\big)\int_0^1\tilde{g}(h,p)dp}. \tag{10}$$

Equation (8) provides a self-contained equation system for the solution of $\tilde{g}(h,p)$, which then directly provides e_h and subsequently $g(p|h)$.

Furthermore, define $g(q,p|h)$ as the fraction of skill-h workers that are matched with productivity-p firms with a bargaining position of q. With the solution for $g(p|h)$ in hand, the steady state condition on $g(q,p|h)$ can, for $q>0$, be written as,

$$g(q,p|h)=\frac{s(h,q)\gamma(p)g(q|h)+s(h,p)\gamma(q)\int_0^q g(q',p|h)dq'}{\delta_0+\hat{\delta}(p)+s(h,p)\widehat{\Gamma}(q)} \tag{11}$$

and, for $q=0$,

$$g(0,p|h)=\frac{\mathbb{1}\left[R(h)\leq p\right]\gamma(p)\left[\frac{1-e_h}{e_h}\left(s_0\left(h\right)+\mu\right)+\delta_0\int_0^1 n(p')\Gamma(p)^{n(p')-1}g(p'|h)dp'\right]}{\delta_0+\hat{\delta}(p)+s(h,p)}. \tag{12}$$

2.5 Recruitment intensity

Firms are modeled as constant returns to scale technologies, each with a single hiring operation where hiring intensity ν is chosen at convex vacancy cost, $\chi\left(\nu\right)$. The output of the firm is the sum of the outputs of each of its matches. The firm meets a worker at rate $\eta\left(\theta\right)\nu$, where $\eta\left(\theta\right)$ is the equilibrium meeting rate of a vacancy. The value of the hiring operation in a type p firm is given by,

$$\begin{aligned}J(p)=\ \max_{\nu}\Bigg[&-\chi(\nu)+\eta\nu(1-\beta)\int_0^1\Bigg[\mathbb{1}\left[R(h)\leq p\right]\left[\eta^u(h)+\eta^0(h,p)\right]\left[V(h,p)-V_0(h)\right]\\&+\int_{R(h)}^p\eta^e(h,p')\left[V(h,p)-V(h,p')\right]dp'\Bigg]dh\Bigg],\end{aligned}$$

where $\eta^u\left(h\right)$ the probability that a meeting is with an unemployed skill-h worker. $\eta^0(h,p)$ is the probability that a meeting is with a skill-h worker coming from an advance notice layoff, and who is willing to match with the firm. Finally, $\eta^e\left(h,p'\right)$ is the probability that a meeting is with a skill-h worker currently employed with a type-p' firm. They are defined by,

$$\begin{aligned}\eta^u(h,p)&=\frac{\psi_h(1-e_h)\big(\kappa s_0(h)+\mu\big)}{\overline{s}},\\\eta^0(h,p)&=\frac{\psi_h e_h\delta_0\int_0^1 n(p')\Gamma(p)^{n(p')-1}g(p'|h)dp'}{\overline{s}},\\\eta^e(h,p)&=\frac{\psi_h e_h s(h,p)g(p|h)}{\overline{s}},\end{aligned}$$

where total worker search, $\overline{s}$, is defined by,

$$\overline{s}=\int_0^1\left[(1-e_h)\big(\kappa s_0\left(h\right)+\mu\big)+e_h\int_0^1\big(s(h,p')+\delta_0 n(p')\big)g(p'|h)dp'\right]\psi(h')dh'. \tag{13}$$

Thus, the optimal vacancy intensity for a type-p firm satisfies the first order condition,

$$\chi'(\nu(p)) = \eta(1-\beta) \int_0^1 \Bigg[\mathbb{1}\left[R(h) \le p\right] \left[\eta^u(h) + \eta^0(h,p)\right] \left[V(h,p) - V_0(h)\right] \\ + \int_{R(h)}^{p} \eta^e(h,q) \left[V(h,p) - V(h,q)\right] dq \Bigg] dh.$$

It is straightforward to show that hiring intensity $\nu(p)$ is increasing in firm productivity since both match value and the acceptance rates are increasing in firm productivity.

2.6 Steady state equilibrium

The sampling distribution from the vacancy pool is the recruitment intensity weighted firm type distribution,

$$\Gamma(p) = \frac{\int_0^p \nu(p')d\Phi(p')}{\int_0^1 \nu(p')d\Phi(p')}. \tag{14}$$

In equilibrium, the meeting rates of workers and firms must balance which implies $\lambda(\theta) = \theta\eta(\theta)$, where, by proportional matching,

$$\theta = \frac{m \int_0^1 \nu(p')d\Phi(p')}{\overline{s}}, \tag{15}$$

where $\overline{s}$ is the total amount of search as defined in equation (13).

With these conditions, the steady state equilibrium can be defined by,

Definition 1. A steady state equilibrium is a collection $\left\{s(h,p), s_0(h), R(h), g(p|h), e, \Gamma(p), \theta\right\}$ that satisfies equations (5), (6), (7), (8), (10), (14), and (15).

2.7 Steady state equilibrium sorting

The steady state equilibrium may or may not display sorting depending on the characteristics of the production function. Our notion of sorting is focused on how worker skill and firm productivity are allocated in steady state equilibrium. We restrict attention to production functions where modularity is global. First, define the worker type conditional CDF of firm types by,

$$\Omega(p|h) = \frac{\int_0^p g(p'|h)dp'}{\int_0^1 g(p'|h)dp'}. \tag{16}$$

Given that search intensity is increasing (decreasing) in worker skill for a supermodular (submodular) production function, following the same logic as in Lentz (2010), the equilibrium is characterized by a strong type of sorting result: For $\kappa \leq 1$, $\Omega(p|h)$ is stochastically increasing (decreasing) in h if and only if the production function is supermodular (submodular).[10]

The firms in our model are multi-worker constant returns to scale firms. Hiring a worker at a given point in time does not preclude future recruitment, so firms are non-discriminatory in hiring. The worker's job acceptance strategy is similarly trivial in a model where employed search is no less efficient than unemployed search. In this case, workers accept any match regardless of firm productivity as long as it is better than the current match.

When the production function is supermodular, more skilled workers have greater gains to upward movement on the firm ladder, and acting on their incentives they consequently search more intensely. Since downward movement on the ladder is skill independent, more skilled workers will—in a stochastic dominance sense—be matched with better firms. In the submodular case, it is the low skill workers that have larger relative gains to upward movement, and so, the low skilled workers search more intensely and end up higher on the firm ladder.

The result generalizes to any $\kappa > 0$ as long as $R(h)$ is increasing (decreasing) in worker skill when the production function is supermodular (submodular). The last qualification is non-trivial. As demonstrated in Shimer and Smith (2000) simple super- or submodularity does not necessarily dictate positive or negative assortative matching (PAM or NAM) through the match acceptance/rejection decision channel. Our model is subject to the same complications. As it turns out, our model estimate implies $\kappa < 1$, and as such, the reservation level is trivially $R(h) = 0$ for all skill levels.

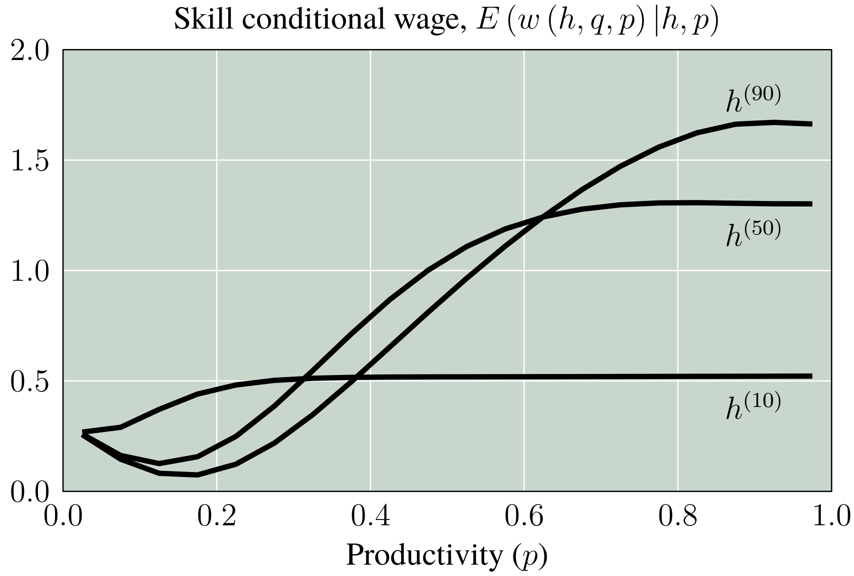
2.8 Wages

Together, equations (1)-(4) determine the wage function $w(h, q, p)$. We provide an analytical expression and interpretation in Appendix D. The worker's current valuation of a match includes an expectation of increasing match value throughout the current employment spell. This is true both within and between jobs in the given employment spell.[11] In particular, total joint match value is increasing in firm type and competition implies that the worker is moving in the direction of extracting this value. In isolation, it implies that the current wage is decreasing in firm type for a given current worker match value, which is why wages may fall as the worker moves between employers.

[10] As in Lentz (2010), the proof (available upon request) establishes the relationship between $s_h(h, p)$ and the worker skill conditional steady state match distribution.

[11] An employment spell is defined as a period of unbroken employment between two unemployment spells, which can include several employers.

Figure 1: Non-monotone wages



Note: The example assumes $\delta(p)=\delta$ and $n(p)=1$ and the parameterization, $(c_0, c_1, \rho, f_0, \alpha, \beta, \kappa, \delta, \delta_0) = (0.01, 1, -7, 5, 0.5, 0.1, 1, 0.1, 0.05)$. In addition, the vacancy and worker skill distributions are assumed uniform. $h^{(x)}$ indicates the xth percentile in the worker skill distribution $\Psi(h)$.

In general, the expected growth rate of match value depends on both worker and firm type. It depends on worker type through the worker's search intensity, which dictates the arrival of outside competition. The firm type enters through this channel as well. In addition, the firm type enters through the layoff rate and the advance notice shock. A decreasing layoff rate in firm type will tend to amplify the wage mechanics as described above in that worker match value growth in low type firms is discounted more heavily than match value growth in high type firms.

Wages are not necessarily monotone in firm type p. This is a well known feature of the sequential bargaining model as in Postel-Vinay and Robin (2002) and Cahuc et al. (2006). As it turns out, wages are also not necessarily monotone in the worker skill index, either. Figure 1 illustrates the average wage steady state wage for an (h, p) match, $\mathbb{E}\big[w(h, q, p) \mid h, p\big]$, where model specification is given in detail in Section 5.1 and parameterization is given in the figure note.

Figure 1 draws the wage as a function of p for the 10th, 50th, and 90th worker skill percentile, denoted $h^{(10)}$, $h^{(50)}$ and $h^{(90)}$. In the example, the production function is supermodular and the worker's bargaining power, β, is relatively low. The example shows that it is possible that worker skill conditional wages may be decreasing in firm type for parts of the firm productivity support. For a given wage, a more productive firm is more valuable to a worker because it offers the possibility of more rent extraction in the event of future outside offers. Hence, for a given bargaining position, negotiations with a more productive firm results in an initially lower wage. However, the realization of future higher wages may tend to take place with an even more productive firm, making it possible that some firms pay lower wages on average than their less productive peers. This is a feature of the wage determination mechanism that does not rely on the modularity of the production function.

Figure 1 also illustrates that it is possible that firm productivity conditional wages may be decreasing in worker type (see for example the 20th percentile firm type in the example). This complete reversal of the ranking of workers by wages stems from supermodularity. The more skilled worker is expecting greater future wage gains relative to a less skilled worker, an effect that is amplified by the greater search effort among more skilled workers when production is supermodular. Consequently, for identical outside options the current wage is lower for the more skilled workers. Using wages across workers within a given firm to identify worker types is further complicated by the possibility that workers outside option q may vary systematically with worker skill type. In particular, in the case of a submodular production technology, low type workers search more intensely and accumulate a better bargaining position. Low skill workers may thus end up with higher wages within a firm.

We emphasize the non-monotonicity of the wage function in the underlying productivity indices because it complicates the worker and firm type classification identification which cannot be done by wages alone. In particular, it takes off the table the otherwise obvious approach of using the wage fixed effect rankings that result from an Abowd et al. (1999) wage decomposition. It also eliminates the identification strategies in Hagedorn et al. (2016) and Bartolucci and Devicienti (2015) where worker classification is obtained subject to wages being monotonically increasing in worker type within firms.

3 Data

Our empirical analysis is conducted using Danish administrative matched employer-employee (MEE) data. The backbone of our data is a population-wide dataset on individual labor market histories recorded at a weekly frequency during 1985-2003. The data comprises information on virtually all individuals residing legally in Denmark aged 15-70. Workers and firms are identified via unique IDs. Individual labor market spells are constructed from administrative registers with information on public transfers, hourly wages, and start and end dates for all jobs reported by employers to tax authorities, and mandatory employer pension contributions.

The raw data identify five labor market states: employment (jobs), unemployment, retirement, self-employment and non-participation. By construction, non-participation is a residual state reflecting that an individual is neither employed nor self-employed nor receiving any kind of public transfer that would categorize him/her as unemployed or retired. Hence, in addition to genuine out-of-the-labor-force spells, non-participation captures imperfect take-up rates of public transfers, reception of transfers not used in the construction of the spell data and misreported start and end dates of spells.

Using person and firm IDs we merge the spells data with information on individual education and wages, and firm's sector of operation from IDA (Integreret Database for Arbejdsmarkedsforskning), an annual population-wide (age 15-70) Danish MEE panel constructed and maintained by Statistics Denmark from several administrative registers. Our wage measure is an estimate of the average hourly wage for jobs that are active in the last week of November. No wage information is available for job spells that do not overlap with a last week of November.[12]

A number of selection criteria and data manipulations are imposed in order to rid the data of invalid observations and to reduce un-modeled heterogeneity as well as other features of the data that our model is not designed to deal with. First, we truncate individual labor market histories at age 55 and discard any labor market history that predates labor market entry as measured by date of graduation from highest completed education. Second, we discard all workers ever observed in employment in the public sector, in self-employment, in retirement, or in the agricultural industry. Third, we recode non-participation spells as unemployment spells.[13] We recode unemployment spells with duration no greater than 13 weeks followed by recall of the worker back to the same employer as part of the original employment spell. In addition, we recode unemployment spells of duration two weeks or less in between two employment spells with different employers as a transition between two job spells within a single employment spell. Fourth, We select the period

[12] In addition, we have information on value added per worker from a subset of firms for the period 1995-2003. This information is obtained from an annual accounting data survey conducted by Statistics Denmark, see Bagger et al. (2014) for further description of this data. We merge value added information onto the labor market history data, but do not utilize it in the estimation. Instead we use it to document how our firm rank measure, a poaching rank to be introduced further below, correlates with observed firm productivity.

[13] Our model features two labor market states and we must decide how to treat the empirical observation of non-participation in relation to the model's notion of unemployment. Coding non-participation as unemployment implies we work with a broad definition of unemployment.

Table 1: Analysis data—Summary statistics

	All years	1994	2003
Number of observations	6,815,884	658,465	703,707
Number of individuals	782,951	552,869	588,643
Number of job spells	1,698,990	490,309	511,604
Number of unemployment spells	608,065	168,155	192,102
Number of firms	117,847	53,537	58,210
Number of firm-years	559,920	53,537	58,249

1994-2003 for our analysis. Our structural model assumes permanent worker and firm types. We want to have a long enough panel to be able to effectively measure worker flows, but do not want to push the type permanence assumption too much. Fifth, we trim the annual individual hourly wage at the 1st and 99th percentiles, and trend them to 2003 levels.

Table 1 provides basic summary statistics on the final analysis panel and also shows statistics for the first (1994) and last (2003) annual cross section in the data. We postpone the presentation of further descriptive statistics to our discussion of model identification in section 4 and the analysis of the structural model's fit in section 5.3.

4 Identification

We estimate the model using Indirect Inference, see Gourieroux et al. (1993). This involves finding the structural parameters that best fits a chosen set of auxiliary statistics. Identification is a question of ensuring that there is one and only one set of structural parameters that is consistent with the auxiliary statistics included in the estimation. In total there are 57 statistics in the auxiliary model. In the following we describe the construction of the statistics and their role in identification.

Identification of labor market sorting requires statistics that rank workers and firms in terms of their unobserved skills and productivities h and p. Such statistics are difficult to find. Wage data alone does not suffice because wages are not monotone in the firm and worker productivity indices, see Section 2.8.[14] Firm level output data does not solve the firm rank identification problem either, as the non-monotonicity extends to labor productivity.[15] Instead, we turn to worker reallocation rate heterogeneity.

Besides the production function, there are many other aspects of our model to be identified. For this reason—and to ensure that the estimated model explains a comprehensive set of labor market outcomes—we include a large number of auxiliary statistics related to worker reallocation and cross section heterogeneity in the estimation. We first describe our firm classification concept which is based on a revealed preference argument.

4.1 The poaching rank

Our model features a firm ladder with higher rungs representing more productive firms. As noted in Section 2.3 total match value, $V(h,p)$, is strictly increasing in p regardless of worker skill. Furthermore, match separation is efficient. Therefore, leaving aside the advance notice shock, any job-to-job transition reflects a revealed preference over the two firms where the poaching firm must have productivity greater than the raided firm. We use this insight to construct a poaching index that allows identification of the productivity rank of firms.[16]

Firm j's poaching index π_j measures the fraction of a firm's hires that are poached from other firms,

$$\pi_j = \frac{N_j^{EE}}{N_j^{UE} + N_j^{EE}}, \tag{17}$$

where N_j^{UE} is the number of firm j hires from unemployment, and N_j^{EE} is the number of workers firm j poached from other firms. Hence, the total number of firm j hires is $N_{\mathrm{j}}^{UE} + N_j^{EE}$. In expectation, a firm's poaching index equals the hire conditional probability that the worker is poached,

[14] Similar conclusions are obtained by Eeckhout and Kircher (2011), Lopes de Melo (2016), and Lise et al. (2016). The mechanisms by which non-monotonicity arise are different in theses papers relative to ours. In Appendix C we document that wage non-monotonicity bias the correlation between worker and firm fixed effects from a log wage regression downward relative to the correlation between h and p. Hence, our analysis suggests that the wage fixed effects correlation provides a lower bound on the true degree of assortative matching.

[15] The equilibrium distribution of h conditional on p, $\Omega_j(h|p)$, $j \in \{L, H\}$, is not necessarily stochastically increasing (decreasing) in p when the production function is supermodular (submodular); see Lentz (2010). Hence, a given firm may have a less skilled workforce than its slightly less productive competitor, implying that a ranking of firms by labor productivity does not necessarily reflect a ranking of firms by underlying productivity p.

[16] Other papers that use job-to-job transitions as revealed preference are Sorkin (2015) and Taber and Vejlin (2016).

which in steady state is given by,

$$\pi(p_j) = \mathbb{E}[\pi_j \mid p_j] = \frac{\int_0^1 \left[\eta^0(h, p_j) + \int_0^{p_j} \eta^e(h, p')dp'\right] d\Psi(h)}{\int_0^1 \left[\eta^u(h, p_j) + \eta^0(h, p_j) + \int_0^{p_j} \eta^e(h, p')dp'\right] d\Psi(h)}, \tag{18}$$

where $\eta^x(h, p)$ for $x \in \{u, e, 0\}$ is defined in Section 2.5. The denominator of equation (18) is the probability that a meeting results in a hire and the numerator is the probability that a meeting results in a poaching. For the case, $\kappa \leq 1$, straightforward differentiation yields the expected poaching index is monotonically increasing in the firm's productivity index p. The probability that a meeting results in a hire from unemployment is unaffected by the firm's type ($\eta_p^u(h, p) = 0$), whereas the probability that a meeting results in a poaching is increasing in the firm's productivity, $\eta_p^0(h, p) + \eta^e(h, p) > 0$.[17]

For each firm in our analysis data, we collect all hires made during the 10 year observation window and record the share of these hires that come directly from other firms, π_j. We proceed to define the firm's poaching rank $\hat{\pi}_j \in [0, 1]$ as firm j's poaching index position in the cumulative poaching index distribution function across the firms. Given that $\mathbb{E}[\pi \mid p]$ is monotonically increasing in p, the poaching rank is an unbiased estimate of the firm's rank in the cumulative productivity distribution function across the firms.

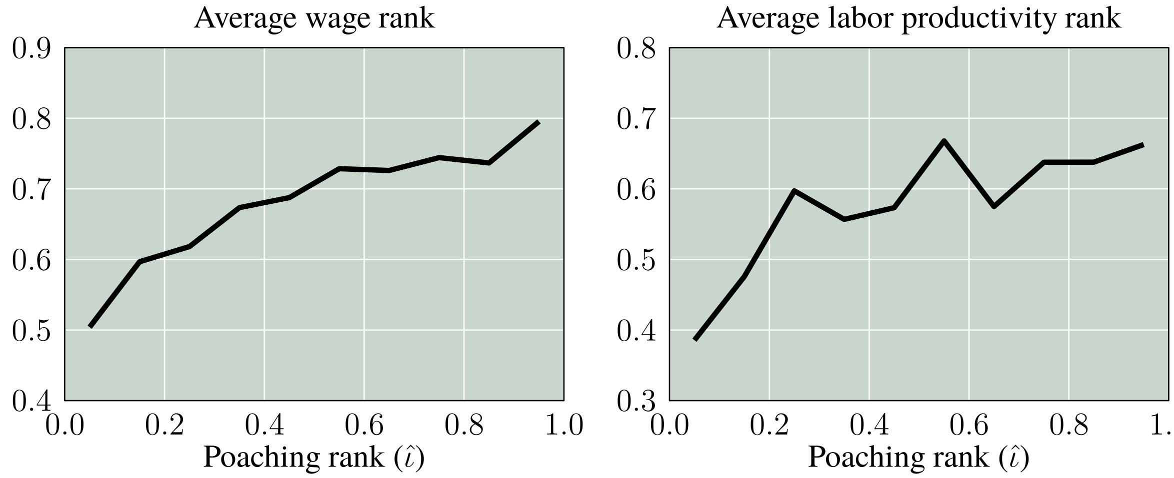
In the estimation, the poaching index is measured only for firms with a total inflow of more than 15 hires, and at least one hire from unemployment. Thus, very small firms are assigned to the group of firms without a rank. Given the prevalence of exogenous job-to-job reallocation in the data, the noise in the inflow measure is considerable for very small firms and conditioning on extreme realizations (either low or high) of the poaching index tends to select very small firms. The poaching index is less noisy for larger firms.[18]

The poaching rank turns out to be closely correlated with other common firm rank measures such as the firm's average wage across its workers or the firm's labor productivity, defined as its value added per worker. In the data, the correlation between the poaching rank and the firm wage rank is 0.36. The correlation between the poaching rank and the labor productivity rank is 0.25. Figure 2 shows the relationship between the poaching rank and the wage and productivity ranks. In contrast

[17] In the case where $\kappa > 1$, the match acceptance decision, as represented by the productivity threshold $R(h)$, must be taken into account. Monotonicity does not necessarily hold in cases where $R(h)$ varies across h. The case $\kappa < 1$ turns out to be the empirically relevant one.

[18] We replicate the same selection in our model simulation, as well as the average firm size in the data so as to emulate the noise in the firm rank measure that is related to small numbers.

Figure 2: Poaching rank conditional average wage and productivity ranks.



Note: The firm wage is its wage bill per full time worker. Firm labor productivity is value added per full time worker.

to both firm wage and labor productivity ranks, the poaching rank is a valid firm classification measure given the structure in the analysis.

4.1.1 Rank conditional mobility

Given the poaching rank, we can investigate the mobility patterns of workers between firms by rank. Specifically, we group firms by rank deciles and calculate worker mobility patterns conditional on the rank of the current firm. The data we use for this exercise is a(n auxiliary) employment spell flow dataset extracted from the analysis data described earlier. Specifically, we select all jobs with non-missing poaching index initiated prior to the final year of our data period. For each job we record duration t, indicators for the job ending in a job-to-job, respectively a job-to-nonemployment transition, and the poaching index $\hat{\pi}$.

The upper, left hand panel of Figure 3 shows the origin rank conditional probability density function of the destination firm's poaching rank given a job-to-job transition. The graph is a contour map and it is seen that the probability mass tend to concentrate above the diagonal. With a few minor imperfections, we find a strong stochastic dominance result: The destination distribution is stochastically increasing in the origin firm rank. This is exactly as one would expect in canonical on-the-job search models such as Burdett and Mortensen (1998) and Postel-Vinay and Robin (2002).[19]

[19] In relation to the upper, left hand panel of Figure 3, it is worth noting that the standard formulation of empirical job ladder models would struggle to explain how the destination rank distribution associated with job-to-job transitions

The upper, right hand panel of Figure 3 shows the average rank of the destination firm conditional on the origin firm rank. Confirming the pattern in the left hand panel, it is seen that, on average and conditional on a job-to-job transition, workers in higher rank firms move to higher rank firms.

The lower left hand panel of Figure 3 shows the rate at which workers move directly from one firm to another. Let $d_{i\tau}^{EE}$ and $d_{i\tau}^{EU}$ be binary indicators for spell i ending in a job-to-job, respectively job-to-unemployment, transition within three months given current quarterly duration τ. We estimate $\widehat{P}_{\tau}^{EE}(\pi) = \Pr(d_{i\tau}^{EE} = 1|\hat{\pi}_i = \pi)$ and $\widehat{P}_{\tau}^{EU}(\pi) = \Pr(d_{i\tau}^{EU} = 1|\hat{\pi}_i = \pi)$ by non-parametrically regressing $d_{i\tau}^{EE}$ and $d_{i\tau}^{EU}$ on $\hat{\pi}_i$. We transform the estimated transition probabilities into hazard rates consistent with the Poisson arrival rates in our model,

$$\vartheta_{\tau}^{EE}(\pi) = -\frac{\widehat{P}_{\tau}^{EE}(\pi)\ln\Big(1 - \widehat{P}_{\tau}^{EE}(\pi) - \widehat{P}_{\tau}^{EU}(\pi)\Big)}{\widehat{P}_{\tau}^{EE}(\pi) + \widehat{P}_{\tau}^{EU}(\pi)},$$

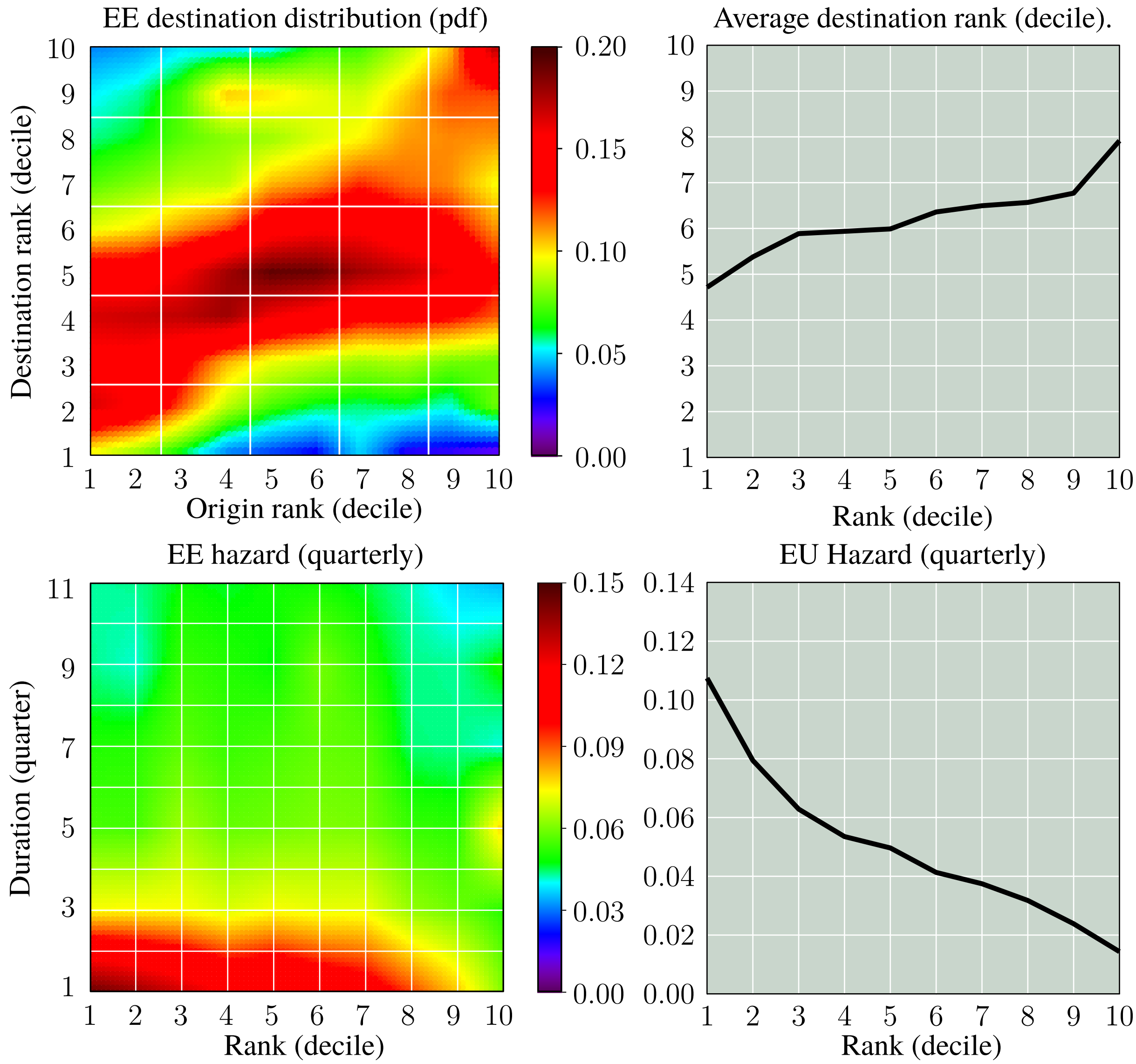
evaluated for 10 deciles in the firm ranking. The figure is a contour map and shows the EE hazard rate conditional on both the rank of the firm and the duration of the match.

The EE hazard rate is strongly decreasing in firm rank for job durations less than a year. For greater job durations, the relationship remains negative but is almost flat. As emphasized in Christensen et al. (2005) the canonical on-the-job search model predicts a decreasing relationship between the EE hazard and firm rank. In higher rank firms, it is less likely that an offer dominates the current match. The search intensity choice amplifies this effect since workers search less intensely for outside offers in higher rank firms. However, the canonical on-the-job search model implies that all workers search at the same intensity within the same firm type. Consequently, the rank conditional EE hazard is unaffected by match duration in these models, as there is no dynamic selection on worker types. The same is true in our model when the production function is modular. Our model interprets the relationship in Figure 3 as a result of variation in search intensity across workers within firms, which occurs when there is sorting.

The lower, right hand panel in Figure 3 shows the first quarter layoff rate into unemployment conditional on firm rank. It is calculated in the same way as the EE hazard. The layoff rate is decreasing in firm rank. The model relies on the firm type conditional layoff rate to match this relationship. As a result it is harder to escape low value labor market states, but once a worker is

out of the highest ranked firms appears to dominate those of lower ranked firms: Job-to-job transitions out of very high ranked firms are mostly involuntary, i.e. forced by "reallocation shocks", which suggests the destination rank distributions from very high ranked firms resemble those of very low ranked firms. We return to this issue when we consider identification and the estimated model's fit, but note here that our advance notice shock specification in fact allow us to reproduce the observed origin rank conditional destination rank distribution pattern.

Figure 3: Job-to-job mobility patterns by poaching rank.

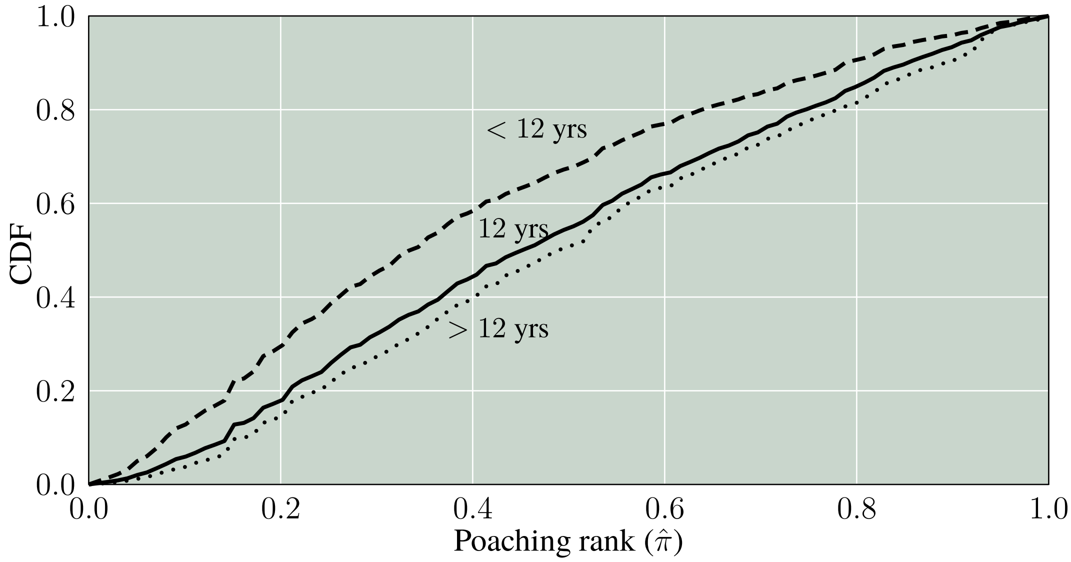


Note: Top left panel: Destination probability is distributed across 10 destination decile bins conditional on origin decile bin. The contour plot illustrates the distributions by interpolation between the 100 points.

in a top firm, there is substantial persistence to the state. This effect is a central feature in Jarosch (2016). In our analysis it is included to ensure that we do not overstate the extent to which workers can escape the lower part of the firm hierarchy through endogenous search.

Finally, in Figure 4 we show the firm poaching rank distribution over matches conditional on the worker's length of education. In the data, 24% of workers have less than 12 years of education. 57% have exactly 12 years of education, and 19% have more than 12 years of education. Figure

Figure 4: Education conditional cumulative firm rank distribution over matches.



Note: Cumulative poaching rank distribution over cross-sectional match distribution conditional on worker's education length. Dashed line conditions on less than 12 years of edcuation. Solid line is for exactly 12 years of education. Dotted line conditions on more than 12 years of edcuation.

4 shows that more educated workers are allocated to higher rank firms. Specifically, the firm rank distribution of a better educated worker stochastically dominates that of a less educated worker. In the estimated model, we find that $\Omega(p|h)$ is stochastically increasing in h. We treat h as unobserved, but it is not unreasonable to expect that worker skill h and education length are positively related. Hence, Figure 4 provides an immediate direct indication that skill and firm productivity are positively correlated in equilibrium, consistent with our model estimate. In additional support of the mechanism that sorting is driven by more skilled workers searching more intensely, Mueller (2010) shows that observed search intensity is increasing in education level in the American Time Use Survey.

4.2 Identifying labor market sorting

We construct two sets of statistics that identify the sorting patterns in the labor market. The first exploits the shape of the firm rank conditional EE hazard rate as shown in Figure 3 where there is less duration dependence in the EE hazard at the top of the firm hierarchy than at the bottom. Our model explains the negative duration dependence in the firm rank conditional EE hazard function through variation in workers' search intensity within firm rank. With this interpretation, the data imply that there is greater search intensity variation in the initial pool of searchers in low rank firms than in high rank firms. Our model naturally delivers this feature when there are complementarities in the production function and consequently sorting. Measurement noise in the firm classification contributes to within-firm rank search intensity heterogeneity, and therefore within-firm rank duration dependence.[20] This measurement noise is replicated in our Indirect Inference estimation procedure and does not bias our estimates.[21] Overall, the firm rank conditional EE hazards dictates some amount of within-firm search intensity variation implying sorting in the estimated model but it is silent about whether sorting is negative or positive.

The second auxiliary statistic related to the identification of production function modularity and labor market sorting is based on the relationship between unemployment durations and reemployment wages in the most productive firms. The empirical correlation between these two variables is negative. The model explains this fact by positive sorting. The following section details the relationship.

4.2.1 Sorting, unemployment durations and reemployment wages

Consider the initial wages of workers that are hired out of unemployment into top rank firms. In the empirically relevant case of $\kappa \leq 1$, these matches are characterized by firms of type $p = 1$ and bargaining position $q = 0$, and the reemployment wage is a direct reflection of the model object,

$$w(h, 0, 1) = \beta f(h, 1) + (1 - \beta) r V_0^*(h), \tag{19}$$

where $rV_0^*(h) = \max_{s \geq 0} \Big\{ f(h, 0) - c(s) + (\mu + \kappa s)\lambda\beta \int_0^1 [V(h, p') - V^0(h)]\, d\Gamma(p') \Big\}$ is the value of unemployment of a skill h worker net of the exogenous reallocation shocks. By equation (19), these particular reemployment wages are monotonically increasing in h, and therefore provide a ranking of workers by skill level. With a supermodular production technology, $s_0'(h) > 0$; that is, unemployed search intensity is increasing in h. That is, more skilled workers have on average shorter unemployment durations. Taken together, supermodular production functions induce a negative correlation between unemployment duration and subsequent initial wages among work-

[20] The empirical poaching ranks contains estimation (sampling) error. Hence, some firms will be misclassified as belonging to a particular decile in the firm productivity distribution. As workers know their employer's true rank, misclassification induces search intensity heterogeneity within a measured firm rank.

[21] Recent notable efforts in firm classification such as Bonhomme, Lamadon, and Manresa (2015) and Lentz, Piyapromdee, and Robin (2016) identify firm classifications in discrete mixture frameworks. It remains an unsolved problem for these analyses that they do not intrinsically understand the uncertainty in the firm classification.

ers hired into top rank firms.[22] Moreover, the stronger the complementarities in the production function, the stronger the correlation between reemployment wages for workers hired out of unemployment by the most productive firms and these workers' unemployment durations. If sorting is negative the correlation has the opposite sign, allowing a distinction between positive and negative sorting.

In terms of empirical implementation, for each worker ever observed in unemployment we compute the individual average unemployment duration, which we denote $\overline{\tau}_i^0$. We use the poaching rank to identify the top rungs of the firm ladder. Specifically, we use the top 5% of firms in the rank distribution. We then select all combinations of individuals and points in time (collected in the set $\widehat{F}^0$) where the individual has just been hired out of unemployment into one of these firms.[23] We find that, the auxiliary statistic of interest is corr$(\overline{\tau}_i^0, w_{it} \mid (i,t) \in \widehat{F}^0)$. In addition, we include the means and standard deviations of $\overline{\tau}_i^0$ and w_{it} for $(i,t) \in \widehat{F}^0$. Empirically, we find corr$(\overline{\tau}_i^0, w_{it} \mid (i,t) \in \widehat{F}^0) = -0.186$. In isolation, this statistic points to positive sorting.

If we condition on education, we find that workers with less than 12 years of education have an average unemployment duration of 76 weeks. The average unemployment duration for workers with exactly 12 years of education is 55 weeks and that of workers with more than 12 years of education is 49 weeks. Thus, more educated workers have shorter durations. In combination with the match allocation evidence in Figure 4 we thus see that more educated workers are matched with higher ranked firms and when they are unemployed they move faster out of unemployment. Subject to a positive correlation between the unobserved skill h and education length we thus see evidence of positive assortative matching as well as the mechanism that generates it, where more skilled workers leave low rank matches faster.

4.3 Worker reallocation

We next present a number of data moments that relate more broadly to worker reallocation that we want our model to reproduce.

[22] If $\kappa > 1$, unemployed search is more efficient than employed search. With a supermodular production function, high skill workers may now reject offers from the bottom of the firm ladder and unemployment durations may be increasing in the worker skill type. As noted above, empirically we find $\kappa \leq 1$.

[23] We measure the starting wage as the wage on record at the first post-transition November cross section date. For the estimation, we reproduce this observation scheme in the simulations.

4.3.1 Hazard rates

To economize on the number of statistics that we fit the model to, we capture the EE hazard characteristics in Figure 3 through the inclusion of the 1st and 7th quarter EE hazards, which helps to identify the search intensity across worker and firm types. We also include the 1st quarter rank conditional EU hazard which identifies the $\delta(p)$ relationship. As noted the EU hazard is strongly decreasing in firm rank and a decreasing layoff rate in firm type will allow the model to explain this relationship.

The job finding rate is identified by $\mathbb{E}[\overline{\tau}_i^0]$ and $\text{sd}[\overline{\tau}_i^0]$. To further discipline the overall relationship between the layoff rates and unemployed search, we include the overall unemployment rate in the data, which is measured at 0.22. We are employing a somewhat broad definition of unemployment which accounts for the relatively high unemployment rate.

4.3.2 Rank conditional destination distribution

We include the origin conditional average destination rank, shown in the upper, right hand panel of Figure 3. The stochastically increasing relationship between the rank of the worker's current firm and the rank of the destination firm in case of a job-to-job transition is consistent with standard on-the-job search models such as Burdett and Mortensen (1998), but taking that model literally, the destination distribution is simply the offer distribution censored to the left at the current firm rank. This of course is not consistent with the data pattern presented in the upper, left panel of Figure 3 where there is non-neglibible probability mass below the diagonal. Our model interprets downward movement on the productivity ladder as a result of advance notice shocks and by measurement noise in the firm classification. The sensitivity of the worker's response to the advance notice as reflected in $n(p)$ is identified by this part of the data.

4.3.3 Employment cycles and unemployment

To quantify the job-to-job reallocation speed relative to the layoff rate into unemployment, we sample employment spells and determine the number of jobs per employment cycle. An employment cycle is a sequence of consecutive job spells with no intervening unemployment spells. Specifically, we extract a sequence of annually stock sampled employment spells. For each cross section of spells, we record the number of jobs in the employment cycle that the job belongs to, counting both left- and right-censored jobs. We include the average number of jobs per employment cycle, and the standard error of the number of jobs per employment cycle in the set of auxiliary statistics that we want to fit. In the data, there are on average 2.182 jobs per employment cycle with a standard deviation of 1.541.

4.3.4 Record statistics

To provide additional information on the advance notice arrival rate δ_0 we consider the probability that a randomly stock sampled job spell ends in a layoff. In an on-the-job search model with constant offer arrival rate and without exogenous job-to-job reallocation, Barlevy and Nagaraja (2013) show that the statistic is bounded below by $1/2$.[24] Using the annually stock sampled employment spells described above, we include in the vector of auxiliary statistics the average share of non-right-censored cross section matches that ends in a job-to-unemployment transition. Empirically, this share is 0.338, inconsistent with a pure on-the-job search model where all job-to-job transitions are from lower to higher ranked firms.

4.4 Wages

We exploit the matched employer-employee wage data to help pin down the worker and firm heterogeneity distributions. The analysis data provide annual measurements on workers' wages for the ongoing jobs in the last week of November which we use to extract an annual matched employer-employee wage panel. The unit of observation is a given (employed) worker in a given year. In this subsection we let i index workers, t index annual cross sections, let $j(i,t)$ be a function that assigns the ID of the employing firm. That is, $j(i,t) = j$ if worker i is employed by firm j in cross section t.

4.4.1 Log wage regression

We include a restricted version of the Abowd et al. (1999) log wage regression in our set of auxiliary models.[25] Specifically, consider the following matched employer-employee panel log wage

[24] Hence, we take information about δ_0 both from this record type statistics as well as from the EE hazard at top rank firms as described above. The use of the latter statistic to identify δ_0 is possibly sensitive to the assumption that all workers agree on the ranking of firms. This is for example not the case in the partnership sorting model in Shimer and Smith (2000) if there are complementarities in the production function. However, the use of the Barlevy and Nagaraja (2013) statistic to inform δ_0 is robust to this issue since it is primarily a reflection of the relative rates by which workers move up their respective ladders.

[25] The full Abowd et al. (1999) log wage regression is not well suited as an auxiliary model due to its computational complexity. However, previewing some of our post-estimation analysis, our model is able to reproduce the Abowd et al. (1999) wage decomposition well.

regression

$$\ln w_{it} = \varphi_{j(i,t)} + \chi_i + \epsilon_{it} \tag{20}$$

where $\varphi_{j(i,t)}$ is a firm effect, χ_i is a worker effect, and ϵ_{it} is a residual. When estimating (20) we impose $E\left[\varphi_{j(i,t)}\epsilon_{it}\right] = E\left[\chi_i\epsilon_{it}\right] = 0$ and $\mathrm{E}\left[\chi_i\varphi_{j(i,t)}\right] = 0$ for all i and t. The first two assumptions impose "exogenous mobility" in the terminology of Abowd et al. (1999), allowing for estimation of the parameters in (20), including the fixed effects, by OLS. The third restriction, not imposed in Abowd et al. (1999), implies uncorrelated firm and worker effects, as estimated from (20). This latter restriction eases the computational burden involved in estimating (20) considerably, allowing us to include statistics based on (20) in the vector of auxiliary statistics. We estimate (20) by OLS and include the average firm effect $\mathbb{E}[\varphi]$ and its standard deviation sd(φ), as well as the standard deviations of the estimated worker effects and the residuals, sd(χ) and sd(ϵ). The average worker effect and the average residual are normalized to zero, so $\overline{\varphi}$ represents the average log wage in the cross section. We find $\mathbb{E}[\varphi] = 5.238$, sd$(\varphi) = 0.179$, sd$(\chi) = 0.218$, and sd$(\epsilon) = 0.134$.

4.4.2 Starting wages and wage growth

To capture the firm ladder effect on wages, we include the first and second moments of the empirical distribution of starting wages in jobs initiated from unemployment. Denote by F^0 the set of (i,t) combinations where worker i has just been hired out of unemployment. Empirically, we find the average log starting wage is $\mathbb{E}[w_{it} \mid (i,t) \in F^0] = 5.09$ with standard deviation sd$[w_{it} \mid (i,t) \in F^0] = 0.29$.[26] Note that these observations are consistent with a firm ladder model where on-the-job search implies the average cross section wage ($\mathbb{E}[\varphi] = 5.238$) exceeds the average starting wage. Furthermore, the average log starting wage for individuals hired directly from unemployment into top firms is, $\mathbb{E}[w_{it} \mid (i,t) \in F^0] = 5.19$. This is dominated by the average log wage in the cross section, which the model understands through the lower initial bargaining position of workers hired directly out of unemployment.

The model links search behavior and within-job wage growth. The quantitative effect of on-the-job search on within-job wage growth depends on workers bargaining power parameter. We add average annual within-job wage growth, which the data reveals to be be 0.011, to the vector of auxiliary statistics.

[26] As noted above we observe the average annual wage of a job. The starting wage is taken to be the average wage during the first calendar year of the job. We of course reproduce this computation in the simulation based estimation procedure.

4.4.3 Mean-min ratio

Hornstein et al. (2011) propose the mean-min ratio as a useful and parsimonious measure of wage dispersion, and argue that a basic wage search model without worker heterogeneity cannot generate enough wage dispersion as reflected in the mean-min ratio. However, the wage process we employ may in fact produce too much dispersion as measured by the mean-min ratio because initial wages can be low. The extent to which this occurs is driven by workers' bargaining power parameter β, and the mean-min ratio thus serves to discipline β in the estimation.[27] The data used for computing the mean-min ratio is the same as that used for estimation of the auxiliary log wage regression. We estimate the minimum wage as the average wage among the lower 5 percentiles in the wage distribution.[28] Denote the estimate minimum wage by $\underline{w}$, and the mean wage by $\overline{w}$. Then we include $Mm = \overline{w}/\underline{w}$ in the vector of auxiliary statistics. The empirical mean-min ratio is 1.85.

5 Model estimation and fit

5.1 Parameterization and estimation

We adopt a search cost function specification where $c\left(\underline{s}\right) = c'\left(\underline{s}\right) = 0$ for some $\underline{s} \geq 0$.[29] The worker's choice of offer arrival rate is in the range $s \in [\underline{s}, \infty[$. This is done to allow the possibility that worker search intensity is not an essential good in the creation of matches. The firm's recruitment cost function is $c_\nu(\nu)$ for recruitment intensity $\nu \in [0, \infty[$. The cost functions are given by increasing and convex functions,

$$c(s) = \frac{\left(c_0(s - \underline{s})\right)^{1+1/c_1}}{1 + 1/c_1} \quad \text{and} \quad c_\nu(\nu) = \frac{\nu^{1+1/c_{\nu 1}}}{1 + 1/c_{\nu 1}}. \tag{21}$$

where $c_0 > 0$ and the recruitment cost function constant has been normalized at unity. $c_1 > 0$ and $c_{\nu 1} > 0$ determine curvatures.

The match production function is specified as a CES function,

$$f(h, p) = f_0\left(\alpha h^\rho + (1 - \alpha)p^\rho\right)^{1/\rho}, \tag{22}$$

[27] In the extreme, if $\beta = 1$ the worker simply gets $w\left(h, p\right) = f\left(h, p\right)$. For lower bargaining power, the initial wage in an employment relationship will be reduced by the expectation of future wage gains.

[28] We use a "raw" version of this statistic rather than on the residual of some Mincerian wage equation controlling for observable heterogeneity.

[29] If $\underline{s}$ is large, it impacts the monotonicity argument of $w(h, 0, 1)$, which in turn modifies the interpretation of the correlation between wages and unemployment duration. For the relevant ranges of $\underline{s}$ this is a secondary issue, and does in any case not affect the validity of the estimation.

where f_0 is a scale parameter, and $\alpha \in [0, 1]$ sets the weight that is put on the skill index relative to the firm productivity index. The modularity of the CES function is governed by ρ. If $\rho < 1$, then the production function is supermodular. It is submodular for $\rho > 1$, and it is modular for $\rho = 1$.

The firm type conditional layoff rate is parameterized by $\delta(p) = \delta_H - (\delta_H - \delta_L)p^{\varsigma}$, where $\delta_H \geq \delta_L$ are the layoff rates for the lowest and highest firm types, respectively. $\varsigma > 0$ governs the interpolation for the firm types in between. With a similar specification, the number of offer draws in case of a layoff shock is specified by, $n(p) = 1 + n_0 p^{\varsigma_0}$, where $n_0 + 1 \geq 1$ is the number of offer draws that workers make in top firms in case of an advance notice shock. By construction, workers make a single draw from the offer distribution when in the lowest firm types. Finally, ς_0 governs the interpolation for the firm types in between.

We parameterize the firm productivity distribution, $\Phi(p)$, as a Beta distribution with parameters $\left(\beta_0^{\Phi}, \beta_1^{\Phi}\right)$ and the worker skill distribution is assumed to be a Beta distribution with parameters $\left(\beta_0^{\Psi}, \beta_1^{\Psi}\right)$.[30] We allow for classical measurement errors ε^w in annual individual wage observations, with $\varepsilon^w \backsim \mathcal{N}(0, \sigma_w^2)$.

The annual discount rate is fixed at $r = 0.05$ and equilibrium market tightness is normalized at $\lambda(\theta) = 1$. The labor force size is normalized at 1. We set $m = 0.091$ matching the number of firms relative to worker population in the data. By targeting the unemployment rate we match the average firm size in the data. Furthermore, c_0 and c_{v0} are not separately identified so we normalize $c_{\nu 0} = 1$. This leaves us with 20 free structural parameters. Let $\boldsymbol{\omega}$ be the structural parameter vector, which we estimate by Indirect Inference (Gourieroux et al., 1993).

The Indirect Inference estimator is

$$\widehat{\boldsymbol{\omega}} = \arg\min_{\omega} [\mathbf{a}(\boldsymbol{\omega}_0) - \mathbf{a}^S(\boldsymbol{\omega})]' \boldsymbol{\Sigma} [\mathbf{a}(\boldsymbol{\omega}_0) - \mathbf{a}^S(\boldsymbol{\omega})],$$

where $\mathbf{a}(\boldsymbol{\omega}_0)$ is a vector of specific auxiliary statistics and data moments computed on real data, a function the true parameter value $\boldsymbol{\omega}_0$, $\mathbf{a}^S(\boldsymbol{\omega}) = \frac{1}{S} \sum_{s=1}^{S} \mathbf{a}_s(\boldsymbol{\omega})$ is the same vector, but computed on S simulated datasets from the structural model at some parameter value $\boldsymbol{\omega}$, and $\boldsymbol{\Sigma}$ is a symmetric and positive definite weighting matrix. In Appendix F, we present estimation results for the efficiently chosen weighting matrix where the variance-covariance matrix of the auxiliary statistics is obtained by bootstrap. Given the size of the data, all the statistics are precisely estimated, but the wage statistics are more so. Therefore, they receive relatively high weight. As it turns out, this results

[30] This means that we are solving for the equilibrium fixed point vacancy offer distribution $\Gamma(p)$ in each simulation iteration.

in a fit to the core reallocation statistics that leaves something to be desired. Given the focus of reallocation in the identification strategy of the paper we will in the main part of the paper present results that achieve a more balanced fit to the auxiliary statistics. We choose a diagonal weighting matrix that puts somewhat less weight on the wage statistics than the efficient weighting matrix does. Together, the two estimates provide a good illustration of identifying power of the reallocation statistics as it pertains to sorting. We report the weighting matrix we employ in Appendix E.

For suitable choices of $\mathbf{a}$, and under regularity conditions, see Gourieroux et al. (1993), $\sqrt{N}(\widehat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0) \to^d \mathcal{N}(\mathbf{0}, (1+S^{-1})[\mathbf{J}'\boldsymbol{\Sigma}\mathbf{J}]^{-1}\mathbf{J}'\boldsymbol{\Sigma}\widehat{\mathbf{W}}\boldsymbol{\Sigma}\mathbf{J}[\mathbf{J}'\boldsymbol{\Sigma}\mathbf{J}]^{-1})$ where N is the number of observations in the data and $\mathbf{J} = \partial \mathbf{a}(\boldsymbol{\omega})/\partial \boldsymbol{\omega}$, evaluated at $\widehat{\boldsymbol{\omega}}$.

5.2 Model estimate

The estimated structural parameters are presented in Table 2. The production function is with an estimate of $\rho = -7.30$ estimated to be supermodular. In section 6 we quantify the strength of the complementarity in terms of the output gains in a perfectly sorted economy. Worker and firm heterogeneity distribution are both estimated to have strictly decreasing densities. In section 7 we provide a wage variance decomposition that quantifies the impact of the estimated heterogeneity on wages.

There is substantial heterogeneity in the layoff rates across firm types. At the very top, the annual layoff rate is only 0.014 whereas in the least productive firm layoffs happen on average every 1.5 years. Forced reallocation through advance notices are estimated to happen relatively infrequently at an annual rate of $\delta_0 = 0.078$, which means on average every 12.8 years. There is at the very top firms a substantially different response to the advance notice shock than at the bottom. At the top, the destination firm is a result of 5.43 draws. Thus, the data firmly reject the simple version of the exogenous reallocation shock where a worker is presented with a single draw from the offer distribution or face unemployment (the Godfather shock). There is substantial curvature in the $n(p)$ function, though, meaning that the deviation from the Godfather shock happens primarily at the very top of the firm distribution (for example, $n(0.9) = 1.7$). The pure Godfather shock implies a bell shaped average destination rank conditional on the rank of the current firm. At the very top, workers move to other firms only in the case of exogenous reallocation, which in the case of the Godfather shock would just be a single draw from the offer distribution. At the bottom of the firm hierarchy, workers accept any outside offer. Hence, the destination distributions out of firms at the

Table 2: Structural parameter estimates

Structural parameter	Estimate	Structural parameter	Estimate
Top layoff rate $\delta(1)$	0.014 (0.001)	$\Phi(p) = Beta(\beta_0^\Phi, \beta_1^\Phi)$	
Bottom layoff rate, $\delta(0)$	0.682 (0.003)	β_0^Φ	0.196 (0.001)
$\delta(p)$ curvature, ς	1.387 (0.003)	β_1^Φ	7.725 (0.011)
Search cost function $\frac{(c_0(s-\underline{s}))^{1+1/c_1}}{1+1/c_1}$		$\Psi(h) = Beta(\beta_0^\Psi, \beta_1^\Psi)$	
Free search, $\underline{s}$	0.010 (0.001)	β_0^Ψ	0.211 (0.002)
c_0	2.606 (0.003)	β_1^Ψ	3.000 (0.005)
c_1	1.475 (0.002)	$f(h,p) = f_0\Big(\alpha h^\rho + (1-\alpha)p^\rho\Big)^{1/\rho}$	
Recruitment cost function $c_\nu(\nu) = \frac{\nu^{1+1/c_{\nu 1}}}{1+1/c_{\nu 1}}$		ρ	-7.303 (0.008)
$c_{\nu 1}$	0.006 (0.001)	α	0.644 (0.002)
Unemployed search efficiency, κ	0.851 (0.002)	f_0	1494.378 (0.023)
Advance notice shock, δ_0	0.078 (0.001)	Workers' bargaining power, β	0.231 (0.001)
Draws at top firm, $n(1)$	5.430 (0.003)	Std. dev., wage measurement error, σ_w	0.052 (0.004)
$n(p)$ curvature, ς_0	19.580 (0.010)		

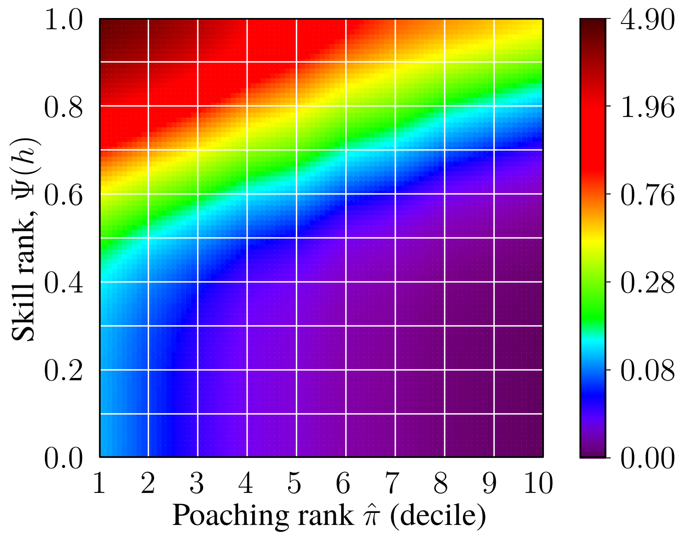
Note: All rates at annual frequency. Standard errors in parentheses.

very top and at the very bottom become identical, which is strongly rejected by the data.

The firm heterogeneity in the layoff and advance notice shocks combine to produce an environment in which there is stronger persistence in the worker's position on the ladder. Workers located on the lower tiers are more likely to be laid off and must climb back up. Workers further up are less likely to be laid off, and when hit with exogenous reallocation shocks, they move subject to stochastically better destination distributions. These mechanics work to amplify the sorting implications from variation in search intensity across workers.

The search cost function is estimated to have somewhat less curvature than a quadratic and there is very little free search. Furthermore, unemployed search is estimated to be less efficient

Figure 5: Search intensity by skill and firm rank.



than employed search, although only by about 15%, $\kappa = 0.85$. Based on survey data on search methods and intensity, Faberman et al. (2016) also find unemployed search to be less efficient than employed search, although significantly more so than we do.

The bargaining parameter is estimated at $\beta = 0.23$. The estimate is broadly consistent with other estimates reported in Cahuc et al. (2006), Bagger et al. (2014) and Bagger et al. (2014). As expected, wages are measured with error although with an estimate $\sigma_w = 0.05$ measurement errors are of modest importance.

Finally, Figure 5 shows the implied search intensities by worker skill conditional on observing the worker in a given rank firm, as measured by the poaching rank $\hat{\pi}$. These skill conditional search intensities aggregate to fit the empirical job-to-job mobility patterns in Figure 3. As dictated by the model, search intensity is decreasing in firm type. Given the supermodular production function, more skilled workers search more than less skilled workers in a given firm rank. Low skill workers are estimated to search relatively little whereas high skilled workers search in low rank firms so as to have spell durations roughly comparable to that of American unemployed workers. It is quite rare in steady state to observe a high skill worker with a low rank firm.

5.3 Model fit

The estimated model's fit to the auxiliary statistics is reported in Table 3. In Appendix F, we show the model fit for the efficiently weighted estimate. As mentioned earlier, the efficiently weighted estimate puts strong emphasis on particular wage related moments at the expense of large parts of the mobility moments. The estimate we are presenting in the main text achieves a more balanced fit. We report the weighting matrix in Appendix E.

5.3.1 Labor market sorting

We have included two sets of statistics that speak directly to the issue of labor market sorting. Moments 28-47 in Table 3 describe the 1st and 7th quarter job-to-job hazards and moment 14 is the correlation between unemployment duration and reemployment wages for workers hired into the top ranked firms.

The model estimate captures nicely the decreasing relationship between current firm rank and the workers propensity to move to other firms. This is a feature emphasized in Christensen et al. (2005). Furthermore, the model does well in fitting that this relationship is stronger in the 1st quarter than in the 7th quarter. In particular, the propensity to leave low rank firms is much stronger in the 1st quarter than it is in the 7th quarter whereas there is little duration dependence in the hazard for the top rank firms. The model explains this through greater heterogeneity in search intensities across workers in low rank firms than in high rank firms, which is associated with sorting.

The model also fits the correlation between unemployment duration and subsequent wages quite well, slightly overshooting it. The model estimate implies a correlation between worker skill and firm productivity in steady state of 0.37. In the efficiently weighted estimate presented in the appendix, the relatively lower weight placed on mobility statistics result in a lower steady state (h, p) correlation of 0.2 which is reflected in a weaker correlation between unemployment duration and reemployment wages (of only -0.07) and the estimate also fails to capture the difference between the 1st and 7th quarter EE hazards in the data. This is a useful illustration of how these moments reveal sorting through worker reallocation.

5.3.2 Worker reallocation

Moments 18-27 in Table 3 are the average destination firm rank conditional on a job-to-job move and the rank of the origin firm. The data show that workers in higher ranked firms tend to move to higher

Table 3: Model fit

	Moment	Data	Model		Moment	Data	Model		Moment	Data	Model
1	$\mathbb{E}[\varphi]$	5.238 (0.0004)	5.268	18	$\mathbb{E}[\iota^d \mid \iota^o = 1]$	4.714 (0.0163)	6.033	38	$\vartheta_7^{EE}(1)$	0.043 (0.0013)	0.070
2	sd$[\varphi]$	0.179 (0.0003)	0.170	19	$\mathbb{E}[\iota^d \mid \iota^o = 2]$	5.378 (0.0163)	6.426	39	$\vartheta_7^{EE}(2)$	0.043 (0.0011)	0.058
3	sd$[\chi]$	0.218 (0.0003)	0.218	20	$\mathbb{E}[\iota^d \mid \iota^o = 3]$	5.887 (0.0142)	6.664	40	$\vartheta_7^{EE}(3)$	0.049 (0.0010)	0.054
4	sd$[\epsilon]$	0.134 (0.0001)	0.165	21	$\mathbb{E}[\iota^d \mid \iota^o = 4]$	5.937 (0.0155)	6.767	41	$\vartheta_7^{EE}(4)$	0.045 (0.0011)	0.050
5	$\mathbb{E}[w_{it} \mid (i,t) \in F^0]$	5.095 (0.0006)	5.047	22	$\mathbb{E}[\iota^d \mid \iota^o = 5]$	5.989 (0.0148)	6.913	42	$\vartheta_7^{EE}(5)$	0.047 (0.0012)	0.044
6	sd$[w_{it} \mid (i,t) \in F^0]$	0.287 (0.0005)	0.202	23	$\mathbb{E}[\iota^d \mid \iota^o = 6]$	6.360 (0.0137)	6.970	43	$\vartheta_7^{EE}(6)$	0.050 (0.0011)	0.041
7	$\mathbb{E}[w_{it} \mid (i,t) \in \widehat{F}^0]$	5.192 (0.0095)	5.130	24	$\mathbb{E}[\iota^d \mid \iota^o = 7]$	6.496 (0.0156)	6.992	44	$\vartheta_7^{EE}(7)$	0.048 (0.0011)	0.038
8	sd$[w_{it} \mid (i,t) \in \widehat{F}^0]$	0.308 (0.0083)	0.222	25	$\mathbb{E}[\iota^d \mid \iota^o = 8]$	6.568 (0.0153)	7.018	45	$\vartheta_7^{EE}(8)$	0.037 (0.0009)	0.035
9	$\mathbb{E}[\Delta w_{it} \mid j(i,t) = j(i,t-1)]$	0.011 (0.0001)	0.008	26	$\mathbb{E}[\iota^d \mid \iota^o = 9]$	6.770 (0.0148)	7.000	46	$\vartheta_7^{EE}(9)$	0.037 (0.0009)	0.032
10	$\overline{w}/\underline{w}$	1.854 (0.0011)	1.615	27	$\mathbb{E}[\iota^d \mid \iota^o = 10]$	7.915 (0.0148)	7.004	47	$\vartheta_7^{EE}(10)$	0.033 (0.0009)	0.030
11	$\mathbb{E}[d_{i\tau}^{EU} = 1 \mid \tau \geq 0]$	0.338 (0.0008)	0.437	28	$\vartheta_1^{EE}(1)$	0.155 (0.0013)	0.193	48	$\vartheta_1^{EU}(1)$	0.107 (0.0011)	0.088
12	$\mathbb{E}[\overline{\tau}_i^0 \mid (i,t) \in \widehat{F}^0]$	1.054 (0.0405)	1.146	29	$\vartheta_1^{EE}(2)$	0.145 (0.0012)	0.152	49	$\vartheta_1^{EU}(2)$	0.079 (0.0010)	0.068
13	sd$[\overline{\tau}_i^0 \mid (i,t) \in \widehat{F}^0]$	1.263 (0.0829)	1.250	30	$\vartheta_1^{EE}(3)$	0.132 (0.0010)	0.129	50	$\vartheta_1^{EU}(3)$	0.063 (0.0008)	0.059
14	corr$(\overline{\tau}_i^0, w_{it} \mid (i,t) \in \widehat{F}^0)$	−0.186 (0.0316)	−0.228	31	$\vartheta_1^{EE}(4)$	0.120 (0.0011)	0.113	51	$\vartheta_1^{EU}(4)$	0.053 (0.0008)	0.052
15	Avg. jobs per emp. spell	2.192 (0.0022)	2.297	32	$\vartheta_1^{EE}(5)$	0.123 (0.0012)	0.090	52	$\vartheta_1^{EU}(5)$	0.050 (0.0008)	0.042
16	Std. dev. job per emp. spell	1.551 (0.0036)	1.571	33	$\vartheta_1^{EE}(6)$	0.114 (0.0011)	0.077	53	$\vartheta_1^{EU}(6)$	0.041 (0.0007)	0.036
17	Unemployment rate	0.218 (0.0004)	0.183	34	$\vartheta_1^{EE}(7)$	0.108 (0.0012)	0.065	54	$\vartheta_1^{EU}(7)$	0.037 (0.0007)	0.031
				35	$\vartheta_1^{EE}(8)$	0.093 (0.0010)	0.054	55	$\vartheta_1^{EU}(8)$	0.032 (0.0006)	0.025
				36	$\vartheta_1^{EE}(9)$	0.077 (0.0009)	0.047	56	$\vartheta_1^{EU}(9)$	0.024 (0.0005)	0.022
				37	$\vartheta_1^{EE}(10)$	0.052 (0.0007)	0.041	57	$\vartheta_1^{EU}(10)$	0.014 (0.0004)	0.019

Note: Standard errors in parentheses (computed by bootstrap). Hazard rates (moments 28-57) are quarterly.

ranked firms conditional on a job-to-job move. The model estimate shows the same pattern but is somewhat flatter than that in the data, overestimating the destination rank distribution for lower rank firms. Hence, the model estimate will tend to underestimate the persistence of a worker's position on the lower part of the firm ladder. The model does however capture well the relationship between firm rank and layoff rate as reflected in moments 48-57 in Table 3, which show the 1st quarter layoff hazards conditional on firm rank. One can through this moment also glean the implied noise in the relationship between a firm's estimated firm rank and its true type. The lowest firm type is estimated to have an annual layoff rate of $\delta(0) = 0.6$ whereas the lowest 10% of firms is estimated to lay off workers at annual rate of roughly 0.35, reflecting some higher type firms have been misclassified. The simulation based estimation method understands this source of noise and the estimate of $\delta(p)$ reflects it.

The estimation matches well the number of jobs in employment cycles (both average and variance, moments 15 and 16 in Table 3). The model overestimates somewhat the probability that a randomly selected spell ends in unemployment, moment 11. This is partly related to the overestimate of the destination distribution of firm types at the low end firms. With a greater propensity to stay in low type firms, these workers would remain subject to the higher layoff rates.[31] Finally, the estimate matches well the unemployment duration distribution of workers entering top firms (moments 12 and 13) as well as the overall unemployment rate in the economy, moment 17. As mentioned earlier, we adopted a somewhat broad definition of unemployment which accounts for both the relatively long average durations as well as the high unemployment rate.

5.3.3 Wage moments

The model fits the log wage decomposition in moments 1-4 in Table 3 very well, slightly overestimating the average wage and the residual wage variance. The model also fits the pattern that the steady state average wage dominates both the initial wage out of unemployment for all workers, moment 5, and the initial wage out of unemployment in top firms, moment 7. Furthermore, the estimation also fits that top firms on average pay more for workers hired directly out of unemployment. It is not trivial to obtain this result in a setting such as Postel-Vinay and Robin (2002) because the greater wage growth rate in top firms makes the initial wage in top firms lower than that of other firms. The lower search intensity in top firms is a helpful feature in our model in this

[31] We conjecture that some degree of direction of search could be a helpful feature in this context, but this is beyond the scope of this paper.

respect. Otherwise, this moment would put a strong discipline on the bargaining parameter, β. The model underestimates the mean-min ratio somewhat but given the nice fit to the overall log wage variance decomposition, we consider it a minor blemish.

Within job wage growth (moment 9 in Table 3) is significant determinant of the bargaining power parameter in the model. The lower the bargaining power, the greater the within job wage growth because accumulated bargaining position becomes a more significant determinant of rent extraction by the worker. Of course, in the extreme where $\beta = 0$ there is no incentive to search and worker reallocation shuts down. The model estimate implies within annual wage growth of 0.8% which is a little below that in the data leaving some room for other accumulation mechanisms that would imply wage growth.

6 Output and mismatch

In a labor market with two-sided heterogeneity, the specific allocation of workers to firms impacts aggregate output. With a supermodular production technology, high skilled workers are more mismatched in low productive firms than low skilled workers. The high skilled workers react by searching intensely for better jobs and therefore leave low productive firms faster than their low skilled colleagues. This endogenous search intensity response alleviates, but does not eliminate, mismatch in the labor market. In this section we quantify the cost of labor market mismatch in terms of foregone output due to mismatch.

Our analysis has two parts. In the first, we assess the output gains from workers' search intensity response to individual mismatch. This analysis involves a comparison between our estimated decentralized economy and two counterfactual planner economies: (a) A planner economy where the planner is unable to induce sorting, and (b) A planner economy where the planner can induce sorting. To remain aligned with the mismatch literature, see e.g. Shimer and Smith (2000) and Gautier and Teulings (2012), where job and worker populations are given exogenously, both planners are constrained to set hiring intensities so as to match the estimated vacancy distribution $\Gamma(\cdot)$, and are also subject to the estimated frictions in the economy. In other words, the planner can set search and hiring intensities, but cannot freely allocate workers across (a fixed set of) vacancies. In the second part, we allow the planner to circumvent labor market frictions and implement a perfectly sorted economy: Given the estimated population of jobs in the economy, the planner directly allocates workers across these jobs in such a way that aggregate output is maximized.

6.1 Search intensity and mismatch

Subject to the vacancy distribution restriction and labor market frictions, a planner may improve on the decentralized economy in two dimensions. First, the total amount of search and recruitment in the decentralized economy may be inefficient. Second, the distribution of search intensities across worker skill levels may be inefficient. The inefficiencies in the decentralized economy arise due to congestion and thick market externalities in the search and recruitment process, see e.g. Hosios (1990). The "sorting planner" can address inefficiency in both these dimensions. The "no sorting planner" can only distinguish between the search intensity of an unemployed worker and an employed worker, but cannot vary search intensity across workers with different skill levels. An output (net of search cost) comparison between the decentralized economy and the sorting planner economy quantifies the total inefficiency in the decentralized economy. A comparison between the sorting and no sorting planner economies quantifies the efficiency gain from sorting through search intensity heterogeneity. Finally, a comparison between the decentralized economy and the no sorting planner economy yields the efficiency gains from pegging the general level of search at the efficient level.

The sorting planner maximizes total output in the economy less vacancy and search costs,

$$\begin{aligned}\mathcal{W} = \max_{s(h,p),s_0(h),\nu(p)} \int_0^1 \Big\{ e_h \int_0^1 \Big[f(h,p) - c\big(s(h,p)\big)\Big] g(p|h) dp \\ + (1-e_h)\Big[f(h,0) - c\big(s_0(h)\big)\Big]\Big\} d\Psi(h) - m \int_0^1 c_\nu\big(\nu(p)\big) d\Phi(p),\end{aligned}$$

where $g(p|h)$ is defined in equation (9).[32] The maximization is done subject to the steady state equilibrium conditions for $\big(g(p|h), e_h, \Gamma(\cdot), \theta\big)$ in equations (9), (10), (14), and (15), as well as the constraint that the estimated vacancy distribution $\Gamma(\cdot)$ is maintained in the planner economies. The no sorting planner can distinguish between unemployed and employed search intensity, (s_0, s, ν), only. In addition, we specify the matching function to be Cobb-Douglas,

$$\lambda(\theta) = A\theta^{\zeta}.$$

Our analysis has not identified the matching function elasticity, ζ. It plays an important role for the planner's ability to substitute vacancy intensity with worker search intensity and vice versa in the creation of matches. Maintaining the estimated vacancy distribution $\Gamma(\cdot)$ in the planner solutions

[32] The expression uses that employed search is at least as efficient as unemployed search, $\kappa < 1$, such that $R(h) = 0$.

Table 4: Decentralized and planner economies

	$\mathcal{W}$	corr(h, p)	Y
Decentralized economy	0.985	0.37	1.000
No sorting planner	0.922	0.00	0.960
Sorting planner	0.999	0.36	1.042
Estimated job and worker populations, perfectly sorted		1.00	1.081

reduces the importance of the parameterization of ζ. Following Petrongolo and Pissarides (2001) we set $\zeta = 0.5$.

The top panel of Table 4 shows the criterion value $\mathcal{W}$, the induced sorting as measured by the correlation between worker skill and firm productivity, corr(h, p), and output, Y, for the three economies described above with the decentralized economy output normalized at unity. The sorting planner improves on the decentralized economy by an increase of output net of search cost of 1.4%. This comes about through a generally higher level of search intensity that reduces unemployment and moves workers higher up the firm ladder while reducing the level of sorting slightly from a correlation coefficient of 0.37 to 0.36. As a result, output increases by 4.2% but the increased search costs offset much of the gain.[33] The no sorting planner solution reported in Table 4 demonstrates the importance of sorting. In this case, where there is no sorting both output as well as output net of search cost drop significantly relative to the decentralized economy. This demonstrates that individual search intensity responses to labor market mismatch provide significant efficiency gains.

6.2 The perfectly sorted economy

The analyses of labor market mismatch in for example Gautier and Teulings (2012), Hagedorn et al. (2016) and in Shimer and Smith (2000) take the populations of jobs and workers as given and ask to what extent the equilibrium in question can be improved upon by changing the allocation of workers to firms. We determine a similar notion of mismatch in our setting: Take the estimated population of jobs in the economy as represented by the employment levels e_h and the distributions $g(p|h)$. Then, assign the estimated population of workers, unemployed and employed, to jobs and unemployment in order maximize aggregate output. We do not concern ourselves with how exactly a social planner might obtain this assignment. The assignment represents a first best non-frictional

[33] The planner solutions are available upon request.

assignment in the spirit of a core assignment in Becker (1973), for a given population of jobs.[34] Since the production function is estimated to be (globally) supermodular, the optimal allocation result in Becker (1973) dictates that the highest skill worker be matched with the most productive job, the second highest skill worker with the second highest productivity job, and so forth.

As shown in the bottom panel of Table 4, we find the perfect allocation of workers to jobs produces an output gain of 8.1%. The result is remarkably close to the measure of mismatch in Gautier and Teulings (2012). While the correlation between h and p in the estimated economy is only 0.37 and therefore suggests substantial misallocation, the magnitude of the efficiency improvement from the perfect allocation depends on the magnitude of the complementarity in the production function over the support of the firm and worker population types. A modular production function would imply a zero efficiency gain from the perfect allocation.

7 Sorting and wage dispersion

We use our structural model to decompose the variance of log wages into worker heterogeneity, firm heterogeneity, imperfect labor market competition, and sorting. We further show that the estimated model is largely consistent with the standard matched employer-employee wage regression pioneered by Abowd et al. (1999), and discuss how our decomposition relate to the log wage variance decompositions presented in Abowd et al. (1999) and Postel-Vinay and Robin (2002) who reach somewhat contrasting conclusions regarding the importance of worker heterogeneity in shaping individual wage outcomes.

7.1 Labor market competition and sorting

The equilibrium wage distribution is a transformation of the equilibrium distribution of worker skill h, firm productivity p, and bargaining position q, formally the productivity of the last firm from which a worker has been able to extract full match value. The distribution of (h, p, q) is given by (11) and (12), and (h, p, q) are mapped into wages w according to (1). Appendix D provides details on the model's wage equation.

The distribution of (h, p, q) is not observed in our data, but is obtained from the estimated structural model. Indeed, it is straightforward to simulate a matched employer-employee dataset

[34] This counterfactual should not be confused with an exercise of eliminating frictions in our model and letting job creation respond in equilibrium.

with individual labor market histories from the model. We simulate the careers of 500,000 workers over 10 years, recording h, p, and q. From this, we compute simulated wages $w = w(h, p, q)$.[35] The object of interest is Var$(\ln w)$.[36]

The implied wage equation is complicated and highly nonlinear in the components h, p, and q, and does not admit an analytical, closed form log wage variance decomposition. To proceed, we base our log wage variance decomposition on predicted log wages from a regression of the simulated log wages $\ln w$ onto indicator variables for the finite number of discrete values for h, p and q. In the estimation, we approximate worker and firm heterogeneity distributions $\Psi(h)$ and $\Phi(p)$ with discrete distributions of 25 distinct h-values and 60 distinct values of p (and hence, 61 distinct values of q). Let $\mathbf{D}_h$, $\mathbf{D}_p$ and $\mathbf{D}_q$ be design matrices of indicator variables for each worker type h, firm type p, and bargaining threshold q in the simulated worker panel. $\mathbf{D}_h$ has generic row $\mathbf{d}'_h$, with $\mathbf{D}_p$ and $\mathbf{D}_q$ constructed analogously. Projecting $\ln \mathbf{w}$ onto the column space of $\mathbf{D} = [1 \ \mathbf{D}_h \quad \mathbf{D}_p \quad \mathbf{D}_q\ ]$—with normalizations to ensure $\mathbf{D}$ has full column rank—yields the (minimum mean square) predicted wages,

$$\ln w^* = \xi_0 + h^* + p^* + q^*. \tag{23}$$

where $h^* = \mathbf{d}'_h \boldsymbol{\xi}_h$, $p^* = \mathbf{d}'_p \boldsymbol{\xi}_p$, $q^* = \mathbf{d}'_q \boldsymbol{\xi}_q$, and $\boldsymbol{\xi} = \left[\ \xi_0 \quad \boldsymbol{\xi}'_h \quad \boldsymbol{\xi}'_p \quad \boldsymbol{\xi}'_q \ \right]' = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' \ln \mathbf{w}$. Unlike the simulated log wages $\ln w$, the predicted simulated wages $\ln w^*$ are linear in a worker skill component h^*, a firm productivity component p^* and a bargaining threshold component q^*. Note, however, that the elements in (h^*, p^*, q^*) are not independent. First, p^* and q^* are mechanically related because $p \geq q$, and because both are increasing in a worker's search capital. Second, labor market sorting implies dependence between h^* and p^* and q^*. The predicted wages, $\ln w^*$, explain 80% of the structural log wage variance, Var$(\ln w)$.

Separating between and within match log wage variance in (23) admits a log wage variance

[35] Observed wages are within-job within-year average wages. In the estimation we appropriately averaged the model's point-in-time wages within-job and within-year, and allowed for measurement errors in observed wages. However, for this analysis of the roles of labor market competition and sorting in shaping the wage distribution, we focus on "pure" cross section wages $w = w(h, p, q)$.

[36] The model is cast in steady state, so the distribution of wages in the panel is identical to the cross section distribution of wages. We conduct the analysis on a simulated panel to increase the number of simulated wage observations and because the panel structure allow us to compare our structural log wage variance decomposition to that obtained from a log wage regressions with worker and firm fixed effects as in Abowd et al. (1999).

decomposition in the mold of Postel-Vinay and Robin (2002).[37] One obtains,

$$\begin{aligned}\mathrm{Var}(\ln w^*) &= \underbrace{\mathrm{Var}(\mathbb{E}[h^* + p^* + q^* \mid p^*, h^*])}_{\text{Between-match}} + \underbrace{\mathbb{E}[\mathrm{Var}(h^* + p^* + q^* \mid p^*, h^*)]}_{\text{Within-match}} \\ &= \underbrace{\mathrm{Var}(h^*)}_{\text{Worker}} + \underbrace{\mathrm{Var}(p^* + \mathbb{E}[q^* \mid p^*, h^*])}_{\text{Firm}} \\ &\quad + \underbrace{\mathbb{E}[\mathrm{Var}(q^* \mid p^*, h^*)]}_{\text{Competition}} + \underbrace{2\mathrm{Cov}(h^*, p^* + \mathbb{E}[q^* \mid p^*, h^*])}_{\text{Sorting}}. \end{aligned} \tag{24}$$

The worker heterogeneity component comprises log wage variation arising from worker skills h^*. The firm heterogeneity component is variation in the "firm effect" $p^* + \mathbb{E}[q^* \mid p^*, h^*]$. Labeling $p^* + \mathbb{E}[q^* \mid p^*, h^*]$ a firm effect is a slight misnomer as $\mathbb{E}[q^* \mid p^*, h^*]$ varies across firm- and worker-types. However, in the absence of sorting, $\mathbb{E}[q^* \mid p^*, h^*] = \mathbb{E}[q^* \mid p^*]$ is a proper firm effect consistent with Postel-Vinay and Robin (2002). The frictional competition reflects within (h^*, p^*)-match bargaining position variation, i.e. it reflects within-match variance in q^*. Variation in q^* is a result of the frictional nature of the labor market. The sorting component arises due to covariance between the worker effect h^* and the firm effect $p^* + \mathbb{E}[q^* \mid p^*, h^*]$. The worker, firm and competition effects are present in the log wage variance decomposition of Postel-Vinay and Robin (2002). The sorting component is new. As the firm effect in (24) has a productivity component, p^*, and a bargaining position component, $\mathbb{E}[q^* \mid p^*, h^*]$, the sorting effect may be further decomposed into sorting on firm productivity, $2\mathrm{Cov}(h^*, p^*)$, and sorting on bargaining position, $2\mathrm{Cov}(h^*, \mathbb{E}[q^* \mid p^*, h^*])$.

The columns labeled "Sorting economy" in Table 5 report the breakdown of log wage variation according to (24) in values and percentage shares of $\ln w^*$. We focus our comments on the percentage shares. Consider first the top panel where we report the decomposition into worker, firm, competition and sorting effects. At 40%, labor market sorting contributes the lion's share of the variance in log wages. Worker heterogeneity comprises 30%, firm heterogeneity 20%, and labor market competition—effectively luck—contributes 10% of the log wage variation. The bottom panel splits the sorting component into sorting on firm productivity, and sorting on bargaining position. Productivity sorting comprises 23%, and sorting on bargaining position 77% of the overall sorting component. Hence, our model estimate implies that more productive workers tend to cluster at more productive firms, and within a given firm, more productive workers command a higher

[37] Postel-Vinay and Robin (2002) derive their wage variance decomposition by separating within- and between-firm variation, but since their model does not allow sorting, covariances between worker and firm types are zero. In our model, with sorting, it is natural to separate within- and between-match variation.

Table 5: Structural log wage variance decomposition: Sorting and no sorting

	Sorting economy		No sorting economy	
	Value	% of Var($\ln w^*$)	Value	% of Var($\ln w^*$)
Worker	0.025	30%	0.016	34%
Firm	0.016	20%	0.012	26%
Competition	0.008	10%	0.019	40%
Sorting	0.033	40%	0.000	0%
Total	0.082	100%	0.047	100%
Sorting				
Productivity, 2Cov(h^*, p^*)	0.003	23%		
Bargaining, 2Cov$(h^*, \mathbb{E}[q^* \mid p^*, h^*])$	0.030	77%		
Total	0.033	100%		

Note: The "No sorting economy" is obtained by setting $n(p) = 1$ for all p and calibrating type-independent unemployed and employed search intensities to match the aggregate unemployment rate ($u = 0.170$) and structural log wage variance ($Var(\ln w) = 0.101$) from the "Sorting economy".

share of the match output through bargaining with the latter effect being quantitatively the more important effect for understanding wage dispersion.

To gain further understanding of the role of sorting in shaping the wage distribution, we compare the structural decomposition (24) at the estimated sorting pattern to the same decomposition obtained in a counterfactual no sorting economy. The no sorting counterfactual also facilitates comparison between our results and Postel-Vinay and Robin (2002). Using French data, they find that worker, firm and competition effects account for 0-35%, 20-50%, and 40-61% of log wage variation, respectively, with shares varying across occupations. We implement the no sorting economy by fixing search intensity levels s_0 and s_1 for unemployed and employed search, respectively, independent of worker skill and employer productivity, as well as restrict the advanced notice shock process to be independent of firm-types, i.e. $n(p) = 1$ for all p. The fixed search intensities s_0 and s_1 are calibrated such that steady state unemployment and variance of log wages are held constant at their estimated levels.[38] We hold all other structural parameters, including the production technology, fixed at their estimated values.

[38]The estimated sorting economy implies a steady state unemployment rate of 0.18 and a structural log wage variance of 0.10. The calibration results in $s_0 = 0.589$ and $s_1 = 1.540$ at annual levels.

The columns labeled "No sorting economy" in Table 5 presents the counterfactual log wage variance decomposition.[39] Eliminating labor market sorting substantially alters the log wage variance decomposition. Without sorting, the variance contributions from worker and firm heterogeneity are stable at 34% and 26% (compared to 30% and 20% in the sorting economy). However, absent sorting, frictional labor market competition emerges as the primary source of wage variation accounting for 40% of log wage variation, up from only 10% with sorting. Despite differences in data sources, estimation procedure, and model specification, the shares reported for the no sorting economy in Table 5 are in line with those reported by Postel-Vinay and Robin (2002) and with the conclusion that the luck associated with frictional labor market competition is a key source of wage dispersion.

Table 5 states that the chance of pure frictional wage dispersion plays a much less prominent role in the sorting economy compared to the no sorting economy. All variation in bargaining position in the Postel-Vinay and Robin (2002) analysis properly falls into this category of chance. However, in the sorted economy the fortune of a good bargaining position is in part the residue of design. Here, more skilled workers search more intensely which results not only in matches with more productive firms, but also in better bargaining positions. This part we attribute to sorting, not luck. Our results are consistent with those of Postel-Vinay and Robin (2002) in that a large portion of wage variance is due to within-match bargaining position variation. However, allowing sorting reveals that a significant part of it is correlated with worker skill which we classify as a sorting contribution.

7.2 Wage regressions and sorting

In their seminal analysis of the distribution of wages, Abowd et al. (1999) pioneered what is now the standard matched employer-employee log wage regression with additive worker and firm fixed effects,

$$\ln w_{in} = \chi_i + \varphi_{\mathrm{J}(i,n)} + \epsilon_{in}, \tag{25}$$

where i index individual workers, n index panel observations (typically, annual cross sections), and $\mathrm{J}(i,n)$ indicates the identity of worker i's employer in cross section n. Hence, χ_i is a worker fixed

[39] Recall that the log wage decompositions are based on (minimum mean square) predicted wages using (23). We estimate separate linear relationships between $\ln w$, h, p and q for the sorting and the no sorting economies. The projection in the no sorting economy has a significantly higher residual term contribution to variance of about 55% whereas it is only 20% in the estimated sorting economy.

effect, $\varphi_{\mathrm{J}(i,n)}$ is a firm fixed effect, and ϵ_{in} represents residual variation. In an actual application, one would typically account for time-varying regressors at both the worker and the firm level. Here, in line with the structural model, we consider the case where all heterogeneity is accounted for by time-invariant worker and firm characteristics that can be subsumed into fixed effects. Abowd et al. (1999) estimate variations and extensions of (25) on French data comparable to that later used by Postel-Vinay and Robin (2002), and find that worker heterogeneity, as measured by variation in the estimated worker fixed effects, is the main driver of log wage variance, although firm effects are also quantitatively important. These results contrast with the those of Postel-Vinay and Robin (2002), who find a limited role for worker heterogeneity in shaping the distribution of wages.[40]

Table 6 presents the log wage variance decomposition obtained from the application of equation (25) to our data as well as simulated data from the estimated model. In Table 6 we denote the correlation between worker and firm fixed effects a "wage sorting effect", not to be confused with the equilibrium correlation between worker skills and firm productivity. The wage sorting effect measures the extent to which high wage workers are matched with high wage firms as defined by the equation (25) fixed effects.

The left two columns in Table 6 present the wage variance decomposition from equation (25) applied to our data. Through this lens worker heterogeneity is the overwhelming contributor to log wage variance: Worker heterogeneity accounts for almost 3/4 of all log wage variance with minor contributions from firm heterogeneity and the residual, and in particular, almost no wage sorting contribution. This is roughly in line with the empirical results presented in Abowd et al. (1999) using French data, but contrasts sharply with our structural log wage variance decomposition where we find sorting to be the single biggest contributor at 40% and worker heterogeneity to account for less than 1/3 of log wage variation.

It is therefore notable that even without targeting these statistics in our estimation, simulated data from our estimated model reproduces the Abowd et al. (1999) log wage variance decomposition quite well (right column in Table 6). In particular, worker heterogeneity is the dominant contributor to variance and the sorting contribution is almost zero. Recall that the estimated model implies a strong equilibrium correlation between worker skill h and firm productivity p at 0.37 with an associated wage variance contribution of 40%. Still, looking at wage data only through the lens of equation (25) we obtain a wage sorting effect of only 6%, which is quite close to its data counterpart.

[40] Postel-Vinay and Robin (2002) contend that their estimates of the contribution of worker heterogeneity to observed wage variation are lower bounds.

Table 6: Log wage variance decomposition—The Abowd et al. (1999) approach.

	Data		Simulated data from estimated model	
	Value	Percent of Var($\ln w_{in}$)	Value	Percent of Var($\ln w_{in}$)
Var($\ln w_{in}$)	0.097	100%	0.107	100%
Worker effect, Var(χ_i)	0.070	72%	0.054	49%
Firm effect, Var($\varphi_{\mathrm{J}(i,n)}$)	0.014	14%	0.023	22%
Residual effect, Var(ϵ_{in})	0.015	16%	0.026	23%
Wage sorting, 2Cov($\chi_i, \varphi_{\mathrm{J}(i,t)}$)	-0.002	-2%	0.007	6%

Hence, we find that a decomposition based on equation (25) significantly biases inference about labor market sorting, and that it overestimates the importance of worker heterogeneity as a source of wage dispersion.

Indeed, the wage equation in our estimated model is not monotone in worker skill and firm productivity which produces negative bias in the wage sorting effect relative to the true sorting contribution, as documented in Appendix C. Moreover, the fixed effects approach of Abowd et al. (1999) is of course not well suited for dealing with the variability in bargaining position arising through on-the-job search. We conjecture that the fixed effects regression of Abowd et al. (1999) attributes the (sizable, see Table 5) within-firm bargaining position variation that correlates with worker skill to the worker wage fixed effect, which is consequently inflated vis-a-vis the worker heterogeneity contribution to the structural log wage variance decomposition in Table 5. Hence, a high wage fixed effect in the Abowd et al. (1999) framework reflects both a high skill-level and a greater ability to extract rent from matches.

7.2.1 Worker wage effect variance by firm type and ID

Our analysis considers sorting of the kind where workers and firms sort by their skill and productivity types. The equilibrium match distribution is estimated to be positively sorted with corr$(h, p) = 0.37$. By implication, the firm type conditional distribution of worker types varies systematically by firm type. In particular, one would expect that the firm type conditional worker type variance be lower than that of the overall worker type population.[41] Thus, an implication of sorting is the result

[41] Eeckhout and Kircher (2011) make related arguments.

Table 7: Worker wage effect sorting by firm type and ID.

	Data	Sim
Var(χ_i)	0.069	0.054
$\mathbb{E}[\text{Var}(\chi_i \mid J(i) = j)]$	0.051	0.051
$\mathbb{E}[\text{Var}(\chi_i \mid \varphi_{J(i)})]$	0.065	0.053

that $\mathbb{E}[\text{Var}(h \mid p)] < \text{Var}(h)$. Indeed, for our estimated model we find that, $\mathbb{E}[\text{Var}(h \mid p)] = 0.012$ while $\text{Var}(h) = 0.015$. The sorting of worker types to firm types induces a sorting of workers within firm types such that there is less difference between workers within firm types than in the overall population. Lopes de Melo (2016) captures this intuition by the correlation between a worker's type and that of his/her co-workers, as measured using Abowd et al. (1999)-type log wage worker fixed effects.

While acknowledging that the wage regression in equation (25) is misspecified and that it is unclear in exactly which way the misspecification biases the fixed effects relative to the underlying worker and firm types, we will nevertheless point to a set of statistics from it that may provide some insight about sorting and avenues for future research.

Consider the cross sectional variance of worker effects, Var(χ_i), reported in Table 7 to be 0.069 in the data and 0.054 in data simulated from the estimated model.[42] In addition, Table 7 shows the worker wage variance within firm and within firm wage fixed effect. We find that the within firm variance, $\mathbb{E}[\text{Var}(\chi_i|J(i) = j)]$, is 0.051 in both data and simulated data from the estimated model. We calculate the firm wage fixed effect conditional variance, $\mathbb{E}[\text{Var}(\chi_i|\varphi_{J(i)})]$ by dividing each cross section into equal size groups of firms with the same firm effects.[43] In the data, we find that $\mathbb{E}[\text{Var}(\chi_i \mid \varphi_{J(i)})]$ is 0.065, while the estimated model implies a within-firm effect variance of 0.053.[44]

As can be seen from Table 7, there is a substantial difference between $\mathbb{E}[\text{Var}(\chi_i|J(i) = j)]$ and $\mathbb{E}[\text{Var}(\chi_i|\varphi_{J(i)})]$ in the data that the model cannot replicate. In the model, all firms with the same productivity have the same expected labor force composition. Thus, by design the model

[42]Note that Var(χ_i) differs slightly from Table 6. The samples differ slightly in that we do not include single worker firms in Table 7.

[43]There are about 50,000 firms in a cross section. We order by firm wage fixed effect and divide into bins each with 50 firms. This procedure corresponds to a variable bandwidth estimator where bandwidth is proportional to the inverse of firm fixed effect density.

[44] For both data and estimate model, $\mathbb{E}[\text{Var}(\chi_i|\varphi_{J(i)})]$ is very close to the unconditional variance, Var(χ_i), leaving little variance in $\mathbb{E}[\chi_i]$ across firm wage effects, consistent with the almost zero wage sorting result.

has $\mathbb{E}[\text{Var}(h \mid p)] = \mathbb{E}[\text{Var}(h_i \mid J(i) = j, p)]$. φ_j is a noisy (and biased) estimate of p_j but it nevertheless introduces only a minor increase in worker wage effect variance within firm types relative to within firms in the estimated model. A similar analysis, omitted here, can be done using Lopes de Melo (2016)'s notion of worker-coworker correlations.

Conditional on firm wage effect, the model explains the data well both in terms of the conditional worker wage variance. This is reassuring in that the model is designed to explain sorting of workers to firms based on the productivity type of the firm. The results demonstrate that sorting of workers to firm productivity type explains a significant reduction of worker type variance within firm relative to the overall population. But the data are clearly suggestive that there is an additional source of sorting of worker groups by firm ID. We conjecture this may reflect a labor market segmentation by worker type. This is beyond the scope of this paper and is not central to its conclusions, but it does suggest that analyses that are specifically directed towards the understanding of the composition of teams within firms would be well advised to consider additional characteristics at the firm level that workers coordinate sorting around.

8 Concluding remarks

A labor market addresses mismatch through worker reallocation. The greater the mismatch, the greater the urgency of the reallocation. Indeed, empirical evidence documents that reallocation is a common occurrence in most labor markets and that reallocation directly between jobs more often than not are associated with wage increases. With an emphasis on mobility patterns in the data, this paper quantifies the contribution of labor market heterogeneity to wage dispersion in a frictional labor market setting where assortative matching may be present.

In the estimated model wage variation is decomposed into four sources: Worker heterogeneity (30%), firm heterogeneity (20%), friction (10%), and sorting (40%). The match production function is estimated to be supermodular implying positive assortative matching. Through the model's wage determination mechanism it incentivizes more skilled workers to search more intensely to reallocate to better firms. The correlation coefficient between worker skill and firm productivity is 0.37 in the steady state match distribution. A social planner can increase output net of search cost by 1.4% which is obtained through higher overall search intensity, but slightly less sorting. In the hypothetical where frictions are eliminated while holding job and worker populations constant, a perfectly sorted economy increases output by 8.1%.

The identification strategy ranks firms by revealed preference through the fraction of a firm's hires that are poached directly from other firms as opposed to hired out of unemployment. We term this the poaching rank. Based on the firm ranking identification is obtained using firm rank conditional mobility and wage patterns. The model matches the data well.

We find that the estimated wage function is non-monotone in worker and firm types. Consequently, we find significant bias in the wage variance decomposition based on the Abowd et al. (1999) approach, which fails to detect any variance contribution from sorting and it attributes too much importance to worker heterogeneity.

Our estimate implies a relatively modest variance contribution from frictional competition of 10%. This is substantially below comparable estimates in Postel-Vinay and Robin (2002) which assumes no sorting. In fact we find comparable levels of wage dispersion that is attributable to variation in bargaining position and as such our model is supportive of the conclusion that frictional sources have a large impact on wage dispersion. But we find that a significant part of the bargaining position variation in the model is related to search choice variation across different skill workers. More skilled workers search harder and tend to have better bargaining positions. We attribute this to the sorting effect and consequently find a significantly smaller role for pure chance.

Appendix

A Match value characteristics, uniqueness and existence.

For a given worker type h and any $p \in [0,1]$, define the mapping,

$$\tilde{V}(h,p) = \max_{s\geq 0, R\geq 0} \frac{\left\{ \begin{array}{c} f(h,p) - c(s) + \lambda s\beta \int_p^1 \tilde{V}(h,p')d\Gamma(p') + [\delta(p) + \lambda\delta_0]\tilde{V}_0(h) \\ +\lambda\delta_0\beta \int_R^1 \tilde{V}(h,p')d\Gamma(p')^{n(p)} \end{array} \right\}}{r + \delta(p) + \lambda\delta_0 + \lambda s\beta\widehat{\Gamma}(p) + \lambda\delta_0\beta\left[1 - \Gamma(R)^{n(p)}\right]} \tag{26}$$

$$\tilde{V}_0(h) = \max_{s\geq 0, R\geq 0} \frac{f(h,0) - c(s) + \lambda(\kappa s + \mu)\beta \int_R^1 \tilde{V}(h,p')d\Gamma(p')}{r + \lambda(\kappa s + \mu)\beta\widehat{\Gamma}(R)}. \tag{27}$$

By Blackwell's sufficient conditions, this is a contraction mapping in $[\tilde{V}(h,p), \tilde{V}_0(h)]$. Thus, there exists a unique fixed point for this mapping. It is straightforward to show that any fixed point of the mapping in equations (2)-(3) must be a fixed point of that of (26)-(27), and vice versa. Hence, there exists a unique solution to equation (2)-(3).

Define by $F(h,V)$ the distribution of match values in the vacancy pool for a type h worker. It is defined by $F(h,V) = \int_0^1 \mathbb{1}[V(h,p) \leq V]d\Gamma(p)$. By the envelope theorem, the derivative of the match value with respect to p can be written as,

$$\begin{aligned} \left(r + \delta(p) + \lambda\delta_0 + \lambda s(h,p)\widehat{\Gamma}(p)\right)V_p(h,p) = f_p(h,p) - \delta_p(p)[V(h,p) - V_0(h)] - \\ \delta_0\beta n_p(p)\int_{R(h)}^{V(h,1)} F(h,V)^{n(p)} \ln F(h,V)dV > 0, \end{aligned}$$

where the inequality follows from $\delta_p \leq 0$ and $n_p \geq 0$ and the presumption that the match in question is viable, that is $V(h,p) \geq V_0(h)$. Hence, match value is monotonically increasing in p.

B Firm size

In steady state, the mass of productivity p firms with n workers $m_n(p)$ must be constant. Hence, the steady state firm size distribution satisfies,

$$0 = \eta(p)\,m_{n-1}(p) + d(p)\,(n+1)m_{n+1}(p) - (\eta(p) + d(p)\,n)\,m_n(p)\,, \tag{28}$$

for all $n \geq 1$ and p. The firm's expected labor force composition is independent of its size. Hence, the expected destruction rate of matches is $d(p)$ for any firm size. Also, in steady state the number of firm births (firms enter with one worker) must equal the number of deaths,

$$\eta(p)m_0(p) = d(p)m_1(p). \tag{29}$$

An alternative interpretation of equation (29) is that firms do not exit, but they just have no economic activity during periods where they have no workers. During such periods they act like potential entrants. In the estimation we do not use entry and exit information from the data, and so, we do not have to take a stand on the issue. Furthermore, it is given that

$$\sum_{n=0}^{\infty} m_n(p) = m\phi(p), \tag{30}$$

where $\phi(p)$ is the firm productivity distribution PDF. Equations (28)-(30) imply that the type conditional firm size distribution $m_n(p)/(m\phi(p))$ is Poisson with arrival rate $\eta(p)/d(p)$,

$$m_n(p) = \left(\frac{\eta(p)}{d(p)}\right)^n \frac{1}{n!} \exp\left(-\frac{\eta(p)}{d(p)}\right) m\phi(p), \tag{31}$$

for all $n \geq 0$.

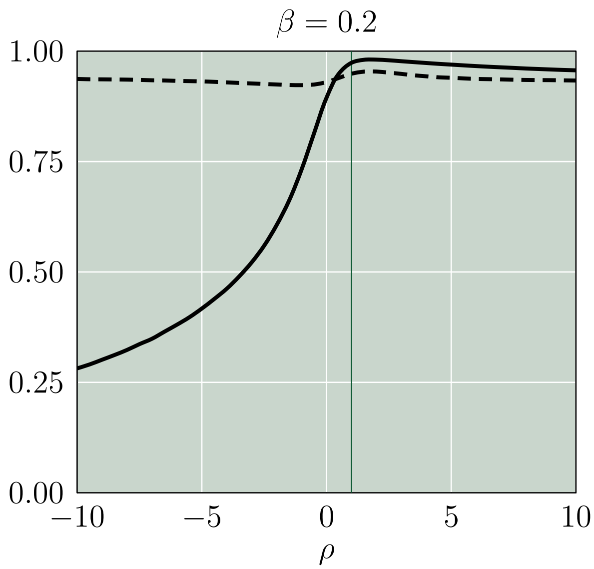
C Wage regressions and monotonicity

Ignoring the role of observable covariates, and subsuming the constant term into, say, the firm effect, Abowd et al. (1999) assume a log wage equation where worker and firm fixed effects enter additively,

$$\ln w_{in} = \chi_i + \varphi_{\mathrm{J}(i,n)} + \epsilon_{in}, \tag{32}$$

where $\mathrm{J}(i,n)$ is the firm ID that worker i is matched with at observation time n, and χ_i and φ_k are the worker and firm fixed effects. The notation as in the main part of the paper, see the description related to our auxiliary log wage regression (20) although here we allow for arbitrary correlation between worker and firm effects. The identification of the fixed effects from matched employer-employee data relies on this additive structure. Consider a class of models where workers differ by skill and firms by productivity. An agent's type is permanent. Furthermore, match output is increasing in both skill and productivity. Can the estimated worker and firm fixed effects from the log-linear wage equation be used as the basis for identification of the underlying worker skill and firm productivity heterogeneity? In particular, does the correlation between the estimated worker and firm fixed effects, corr$(\widehat{\chi}_i, \widehat{\varphi}_{\mathrm{J}(i,n)})$, identify sorting in the matching between worker skill and firm productivity? Eeckhout and Kircher (2011) provide a negative answer for their model. We will generally provide a negative answer as well. Both answers are based on the insight that for the model structures in question, the log additive wage equation is fundamentally misspecified with respect

Figure 6: The correlation between wage fixed effects and true agent heterogeneity for given (ρ, β) combinations.



Note: The solid and dashed lines show cor$[\hat{\chi}, h]$ and cor$[\hat{\varphi}, p]$, respectively. For the given model specification, the production function scale parameter (f_0) and the base offer arrival rate (λ) are set such that the the steady state equilibrium solution satisfies $u = 0.05$ and $\mathrm{E}[w(h, p)] = 180.0$. The dashed red line at $\rho = 1$ divides the model specifications with positive sorting for $\rho < 1$ and negative sorting for $\rho > 1$.

to the worker and firm heterogeneity contributions to wages. Specifically, wages are generally not monotonically increasing in skill and productivity.

In Figures 6 and 7 we relate estimates of worker and firm fixed effects from the wage equation (32) to the true underlying worker skill and firm productivity heterogeneity in simulations of steady state equilibria for different (ρ, β) combinations.

Figure 6 shows corr$(\widehat{\chi}_i, h_i)$ and corr$(\widehat{\varphi}_k, p_k)$. It is seen that the wage equation firm fixed effect is strongly correlated with firm productivity regardless of the type and strength of sorting and worker's bargaining power. Not surprisingly, higher bargaining power increases the correlation.

The correlation between the wage equation worker fixed effect and worker skill is on the other hand quite sensitive to the specification of the model. If sorting is positive and wage determination is primarily set by wage posting, then the correlation is low. In this case, the wage profiles of more skilled workers are characterized by substantial wage growth over an employment spell, and consequently, the notion of a wage equation worker fixed effect is misplaced. As documented in Figure 1 it is in this type of equilibrium also perfectly possible to observe more skilled workers receive lower wages than less skilled workers within a given firm. In such a case, the estimation will tend to rank the less skilled worker with a higher fixed effect than the more skilled worker. This mechanism is strengthened by the assumption that the wage equation has an i.i.d. over time error process, ϵ_{in} and the fact that even for the high skilled workers, the wage process has some permanence to it. Since the more skilled worker's realized wage growth is often associated with an actual job-to-job transition, the estimation will be allowed to explain the substantial observed wage growth of the high skilled worker by increasing the wage equation fixed effect differential between the two firms involved in the job-to-job transition, thereby laying a foundation for a negative bias in the correlation between wage equation worker and firm fixed effects.

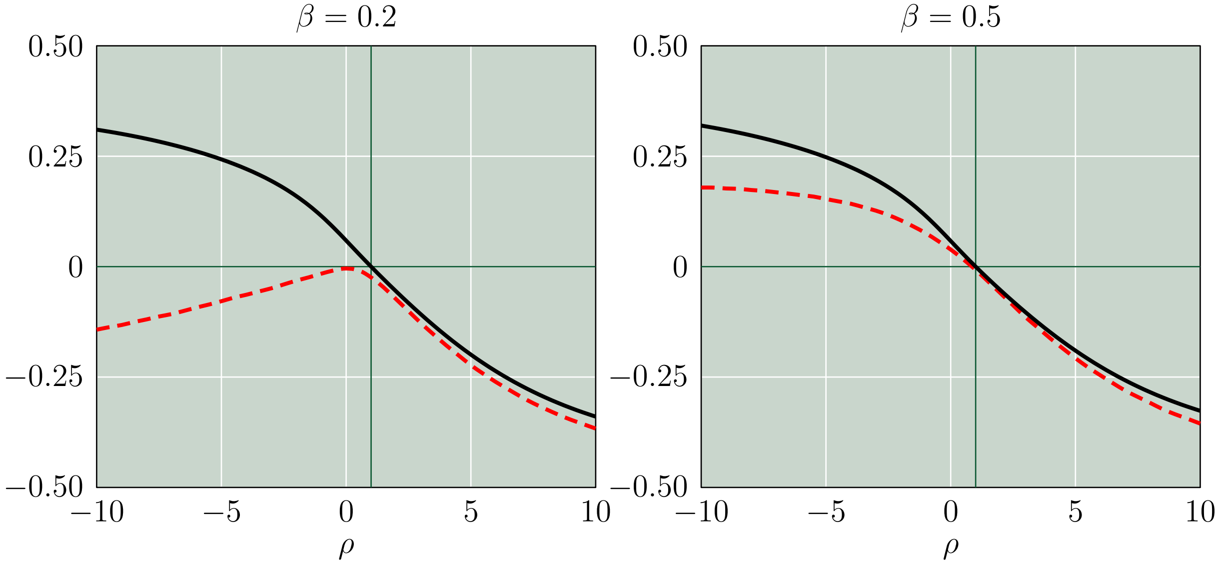
In the negative sorting case, low skilled workers are the ones taking temporary current wage hits with the expectation of future gains. As a result, in this type of equilibrium wages are monotonically increasing in worker skill within a given firm and the ranking of wage equation worker fixed effects will be aligned with the skill ranking. This accounts for the strong positive correlation between the estimated wage equation worker fixed effects and worker skill for the negative sorting cases, $\rho > 1$.

For higher β, where wage determination is to a greater extent set by bargaining rather than posting, $\text{corr}(\widehat{\chi}_i, h_i)$ is higher because wages are moving towards being monotone in worker skill and firm productivity.

Figure 7 presents the correlation between the wage equation fixed effects in relation to the correlation between the skill and productivity indices in the equilibrium steady state match distribution. The correlation between h and p based on $G(h, p)$ reveals the basic property of the model that sorting is positive for $\rho < 1$, negative for $\rho > 1$, and there is no sorting when $\rho = 1$. It is seen that when $\beta = 0.2$ and there is negative sorting, the correlation between wage equation worker and firm fixed effects, $E_n[\text{corr}(\widehat{\chi}_i, \widehat{\varphi}_{\text{J}(i,n)})]$ is very close to equilibrium steady state $\text{corr}(h, p)$. This is consistent with the results in Figure 7 that the estimated wage equation worker and firm fixed effects are closely correlated with the skill and productivity indices in this case. When sorting is positive and $\beta = 0.2$, we see that $E_n[\text{corr}(\widehat{\chi}_i, \widehat{\varphi}_{\text{J}(i,n)})]$ and $\text{corr}(h, p)$ diverge. In this case, the worker fixed effects are so poorly related to the skill ranking that the resulting negative bias drives the correlation between χ and φ negative. As a result, $E_n[\text{corr}(\widehat{\chi}_i, \widehat{\varphi}_{\text{J}(i,n)})]$ is negative both when sorting is positive and negative for this case.

In the case where $\beta = 0.5$, the fixed effects correlation $E_n[\text{corr}(\widehat{\chi}_i, \widehat{\varphi}_{\text{J}(i,n)})]$ does quite well in capturing the steady state match correlation between skill and productivity. There is some negative bias in the positive sorting case, but in this case, the correlation coefficients share the same signs.

Figure 7: The correlation between skill and productivity for given (ρ, β) combinations.



Note: The solid line is cor$[h, p]$. The dashed line is cor$[\chi_i, \varphi_{\mathrm{K}(i,n)}]$. The wage equation fixed effects are estimated on simulated data from the given steady state equilibrium. For the given model specification, the production function scale parameter (f_0) and the base offer arrival rate (λ) are set such that the the steady state equilibrium solution satisfies $u = 0.05$ and $\mathrm{E}[w(h,p)] = 180.0$. The dashed red line at $\rho = 1$ divides the model specifications with positive sorting for $\rho < 1$ and negative sorting for $\rho > 1$.

The above results suggest that an observed positive value of $E_n[\mathrm{corr}(\widehat{\chi}_i, \widehat{\varphi}_{\mathrm{J}(i,n)})]$ indicates that sorting between skill and productivity is positive. In general, the correlation coefficient between h and p is always greater than $E_n[\mathrm{corr}(\widehat{\chi}_i, \widehat{\varphi}_{\mathrm{J}(i,n)})]$. It is also worth emphasizing that the often observed small and negative correlation between χ and φ is consistent with anything from mild negative sorting to strong positive sorting between h and p.

D The wage equation

As described in section 2.8, equations (1), (2), and (4) determine the wage function $w(h, q, p)$ describing the wage offered to a type-h worker by a type-p firm competing for the worker's services with a firm of type $q \leq p$. The wage equation takes the form

$$
\begin{aligned}
w - c(s(h,p)) \;=\;& \underbrace{\beta\Big[f(h,p) - c(s(h,p))\Big] + (1-\beta)\Delta(p,q)\Big[f(h,q) - c(s(h,q))\Big]}_{\text{Net of search cost output sharing}} \\
& -\Big[\underbrace{\lambda s(h,p)(1-\beta)\int_q^p V_p(h,p')\hat{\Gamma}(p')dp'}_{\text{Offer matching, within-firm}} + \underbrace{(1-\beta)\lambda s(h,p)\beta\int_p^1 V_p(h,p')\hat{\Gamma}(p')dp'}_{\text{Offer matching, between-firm}}\Big] \\
& \underbrace{+\lambda s(h,q)(1-\beta)\Delta(p,q)\beta\int_q^1 V_p(h,p')\hat{\Gamma}(p')dp'}_{\text{Offer matching, outside option firm}} \\
& \underbrace{+(1-\beta)\Big[\Delta(p,q)\delta(q) - \delta(p)\Big]V_0(h)}_{\text{Layoff rate heterogeneity (direct effect)}} \\
& \underbrace{+(1-\beta)\lambda\delta_0\beta\int_{R(h)}^1 V_p(h,p')\Big\{\Delta(p,q)\left[1-\Gamma(p')^{n(q)}\right] - \left[1-\Gamma(p')^{n(p)}\right]\Big\}}_{\text{Advance notice shock heterogeneity (direct effect)}}, \qquad (33)
\end{aligned}
$$

where $\Delta(p,q) := [r+\delta(p)+\lambda\delta_0][r+\delta(q)+\lambda\delta_0]^{-1} < 1$ reflects that income flows are discounted differently in type-p and type-q matches due to layoff rate differences.

Wages net of search costs reflect the sharing of match output net of search cost, $f(h,p) - c(s(h,p))$ given the worker's appropriately discounted outside option of employment in the type-q firm, with adjustments for differences in expected future wage trajectories between the type-p and the type-q firms in relation to climbing the productivity ladder, advance notice shocks, and layoff shocks.

First, the term labeled "Offer matching, within-firm" reflects the more productive type-p employers ability to match outside offers, generating within-job wage growth. Second, the term "Offer matching, between-firm" reflects that the more productive type-p firm acts as a better outside option in future wage negotiations with even more productive firm types. Notice that the within- and between-firm offer matching terms are negative: More productive firms offer more rewarding future wage trajectories and use this to depress current wages while remaining competitive vis-a-vis the less productive type-q firm.

Second, while the type-p firm is better able to match future offers, future offers arrive faster in the type-q firm due to $s(h,q) \geq s(h,p)$. A type-h worker negotiating a contract with a type-p firm using a type-q firm as outside option understands this and must be compensated for giving up the faster arrival of outside offers at the type-q firm. This limits the ability of the type-p firm backload wages while remaining competive in hiring and retention. The term labeled "Offer matching, outside option firm" captures this effect. Notice that this term is positive. Further notice that offer matching at the outside firm has wage implications only because of job offer arrival rate heterogeneity.

Third, the term labeled "Layoff rate heterogeneity (direct effect)" captures that jobs in less productive firms are less secure. The larger the difference between $\delta(q)$, appropriately weighted to with the relative discount rate $\Delta(p,q)$, and $\delta(p)$, the more weight is put on the value of unemployment in the wage setting procedure. Layoff rate heterogeneity also has indirect wage effects running through workers' job search behavior and the value of climbing the productivity ladder. Lower layoff rates at the top of the productivity ladder increases the value of jobs at the top rungs, enabling top-rung firms to depress the wage they offer while retaining the ability to hire workers from less productive firms.

Fourth, the term labeled "Advance notice shock heterogeneity (direct effect)" comprises the (direct) wage effects arising from the fact that the distribution from which an alternative job offer is drawn is better (in the sense of stochastic dominance) in the more productive type-p firm. This allows the more productive type-p firm to lower the wage while remaining competitive vis-a-vis the less productive type-q firm. That is, the direct effect of advance notice shocks is negative. Of course, advance notice shock heterogeneity also has indirect effects operating through job search. Indeed, advance notice shock heterogeneity alters the returns to climbing the productivity ladder, and the wage effects of this depends on the $n(p)$-profile, i.e. how the sampling distribution associated with an advance notice shock varies across the productivity ladder.

Our wage equation (33) has a number of interesting special cases. First, for $\beta = 1$, $w = f(h,p)$. In this case the worker appropriates the entire match value and wages simply reflect the match productivity.[45] An identical wage implication arises if $q = p$, i.e. when the worker has achieved full surplus extraction at the current employer. Second, consider the case where there is no advance notice shock heterogeneity, i.e. $n(p) = 1$ for all p, no layoff rate heterogeneity, $\delta(p) = \delta$ for all p, and where search intensity is independent of worker skill h and firm productivity p, i.e. $s(h,p) = s$ for all h, p. This environment does not feature sorting, and was studied by Postel-Vinay and Robin (2002) and Cahuc et al. (2006). Under these restrictions, it is straightforward to show that

$$w = \beta f(h,p) + (1-\beta) f(h,q) - (1-\beta)^2 \lambda s \int_q^p \frac{f_p(h,p')\hat{\Gamma}(p')dp'}{r + \delta + \delta_0\lambda + \beta\lambda s\hat{\Gamma}(p')}$$

[45]The case $\beta = 0$ is of little interest when search intensity is endogenous as it leaves workers with no incentives to search.

where the wage offered to a type-h worker by a type-p firm in competition with a less productive type-q firm reflects output sharing adjusted with an inter-temporal transfer accounting for the ability of the type-p firm to match future offers in the range$p' \in (q, p]$. Furthermore, if match output is multiplicatively separable in h and p, wages are proportional to the worker skill component. This result does not carry over to the case of endogenous search intensities, where worker skill level h enters indirectly the through the search intensity choice.

E Weighting matrix used in main estimation

Table 8: Diagonal elements of weight matrix used in main estimation

		Moment value	Diagonal element			Moment value	Diagonal element
1	$\mathbb{E}[\varphi]$	5.238	0.091	29	$\vartheta_1^{EE}(1)$	0.155	1.062
2	sd$[\varphi]$	0.179	2.665	30	$\vartheta_1^{EE}(2)$	0.145	0.531
3	sd$[\chi]$	0.218	2.180	31	$\vartheta_1^{EE}(3)$	0.132	0.531
4	sd$[\epsilon]$	0.134	3.543	32	$\vartheta_1^{EE}(4)$	0.120	0.531
5	$\mathbb{E}[w_{it} \mid (i,t) \in F^0]$	5.095	0.012	33	$\vartheta_1^{EE}(5)$	0.123	0.531
6	sd$[w_{it} \mid (i,t) \in F^0]$	0.287	0.207	34	$\vartheta_1^{EE}(6)$	0.114	0.531
7	$\mathbb{E}[w_{it} \mid (i,t) \in \widehat{F}^0]$	5.192	0.011	35	$\vartheta_1^{EE}(7)$	0.108	0.531
8	sd$[w_{it} \mid (i,t) \in \widehat{F}^0]$	0.308	0.193	36	$\vartheta_1^{EE}(8)$	0.093	0.531
9	$\mathbb{E}[\Delta w_{it} \mid j(i,t) = j(i,t-1)]$	0.011	5.276	37	$\vartheta_1^{EE}(9)$	0.077	0.531
10	$\overline{w}/\underline{w}$	1.854	0.032	38	$\vartheta_1^{EE}(10)$	0.052	1.062
11	$\mathbb{E}[d_{i\tau}^{EU} = 1 \mid \tau \geq 0]$	0.338	0.176	39	$\vartheta_7^{EE}(1)$	0.043	1.062
12	$\mathbb{E}[\overline{\tau}_i^0 \mid (i,t) \in \widehat{F}^0]$	1.054	0.056	40	$\vartheta_7^{EE}(2)$	0.043	0.531
13	sd$[\overline{\tau}_i^0 \mid (i,t) \in \widehat{F}^0]$	1.263	0.047	41	$\vartheta_7^{EE}(3)$	0.049	0.531
14	corr$(\overline{\tau}_i^0, w_{it} \mid (i,t) \in \widehat{F}^0)$	-0.186	0.639	42	$\vartheta_7^{EE}(4)$	0.045	0.531
15	Avg. jobs per emp. spell	2.192	0.027	43	$\vartheta_7^{EE}(5)$	0.047	0.531
16	Std. dev. job per emp. spell	1.551	0.038	44	$\vartheta_7^{EE}(6)$	0.050	0.531
17	Unemployment rate	0.218	0.183	45	$\vartheta_7^{EE}(7)$	0.048	0.531
18	$\mathbb{E}[\iota^d \mid \iota^o = 1]$	4.714	0.192	46	$\vartheta_7^{EE}(8)$	0.037	0.531
19	$\mathbb{E}[\iota^d \mid \iota^o = 2]$	5.378	0.096	47	$\vartheta_7^{EE}(9)$	0.037	0.531
20	$\mathbb{E}[\iota^d \mid \iota^o = 3]$	5.887	0.096	48	$\vartheta_7^{EE}(10)$	0.033	1.062
21	$\mathbb{E}[\iota^d \mid \iota^o = 4]$	5.937	0.096	49	$\vartheta_1^{EU}(1)$	0.107	1.062
22	$\mathbb{E}[\iota^d \mid \iota^o = 5]$	5.989	0.096	50	$\vartheta_1^{EU}(2)$	0.079	0.531
23	$\mathbb{E}[\iota^d \mid \iota^o = 6]$	6.360	0.096	51	$\vartheta_1^{EU}(3)$	0.063	0.531
24	$\mathbb{E}[\iota^d \mid \iota^o = 7]$	6.496	0.096	52	$\vartheta_1^{EU}(4)$	0.053	0.531
25	$\mathbb{E}[\iota^d \mid \iota^o = 8]$	6.568	0.096	53	$\vartheta_1^{EU}(5)$	0.050	0.531
26	$\mathbb{E}[\iota^d \mid \iota^o = 9]$	6.770	0.096	54	$\vartheta_1^{EU}(6)$	0.041	0.531
27	$\mathbb{E}[\iota^d \mid \iota^o = 10]$	7.915	0.192	55	$\vartheta_1^{EU}(7)$	0.037	0.531
				56	$\vartheta_1^{EU}(8)$	0.032	0.531
				57	$\vartheta_1^{EU}(9)$	0.024	0.531
				58	$\vartheta_1^{EU}(10)$	0.014	1.062

F Efficiently weighted estimate

Here we present the efficiently weighted estimate. We obtain the variance-covariance matrix of the auxiliary model statistics through bootstrap of worker histories with resampling. The weighting matrix is the inverse of the variance-covariance matrix. We provide the parameter estimates without estimated errors.

Table 9: Structural parameter estimates (efficiently weighted).

Structural parameter	Estimate	Structural parameter	Estimate
Top layoff rate $\delta(1)$	0.011 (0.0001)	$\Phi(p) = Beta(\beta_0^\Phi, \beta_1^\Phi)$	
Bottom layoff rate, $\delta(0)$	0.585 (0.0005)	β_0^Φ	0.195 (0.0004)
$\delta(p)$ curvature, ς	1.629 (0.001)	β_1^Φ	6.972 (0.001)
Search cost function $\frac{(c_0(s-\underline{s}))^{1+1/c_1}}{1+1/c_1}$		$\Psi(h) = Beta(\beta_0^\Psi, \beta_1^\Psi)$	
Free search, $\underline{s}$	0.003 (0.0003)	β_0^Ψ	0.252 (0.0004)
c_0	3.318 (0.001)	β_1^Ψ	2.329 (0.001)
c_1	0.647 (0.0005)	$f(h,p) = f_0\Big(\alpha h^\rho + (1-\alpha)p^\rho\Big)^{1/\rho}$	
Recruitment cost function $c_\nu(\nu) = \frac{\nu^{1+1/c_{\nu 1}}}{1+1/c_{\nu 1}}$		ρ	-7.563 (0.001)
$c_{\nu 1}$	0.005 (0.0003)	α	0.436 (0.0005)
Unemployed search efficiency, κ	0.590 (0.0004)	f_0	1492.961 (0.005)
Advance notice shock, δ_0	0.084 (0.0002)	Workers' bargaining power, β	0.352 (0.0004)
Draws at top firm, $n(1)$	2.095 (0.001)	Std. dev., wage measurement error, σ_w	0.010 (0.004)
$n(p)$ curvature, ς_0	19.652 (0.003)		

Note: All rates at annual frequency. Standard errors in parentheses.

The fit is given in Table 10. As can be seen, the sorting moments are matched less well and point to less sorting, which is indeed born out in a steady state distribution correlation between h and p of 0.19. Given the focus on the mobility part of the model, we chose to given additional weight to these moments in the main estimate. As mentioned it is however a useful illustration that

Table 10: Model fit

	Moment	Data	Model		Moment	Data	Model		Moment	Data	Model
1	$\mathbb{E}[\varphi]$	5.238 (0.0004)	5.238	18	$\mathbb{E}[\iota^d \mid \iota^o = 1]$	4.714 (0.0163)	6.069	38	$\vartheta_7^{EE}(1)$	0.043 (0.0013)	0.095
2	sd$[\varphi]$	0.179 (0.0003)	0.192	19	$\mathbb{E}[\iota^d \mid \iota^o = 2]$	5.378 (0.0163)	6.388	39	$\vartheta_7^{EE}(2)$	0.043 (0.0011)	0.072
3	sd$[\chi]$	0.218 (0.0003)	0.194	20	$\mathbb{E}[\iota^d \mid \iota^o = 3]$	5.887 (0.0142)	6.530	40	$\vartheta_7^{EE}(3)$	0.049 (0.0010)	0.062
4	sd$[\epsilon]$	0.134 (0.0001)	0.142	21	$\mathbb{E}[\iota^d \mid \iota^o = 4]$	5.937 (0.0155)	6.603	41	$\vartheta_7^{EE}(4)$	0.045 (0.0011)	0.055
5	$\mathbb{E}[w_{it} \mid (i,t) \in F^0]$	5.095 (0.0006)	4.963	22	$\mathbb{E}[\iota^d \mid \iota^o = 5]$	5.989 (0.0148)	6.672	42	$\vartheta_7^{EE}(5)$	0.047 (0.0012)	0.046
6	sd$[w_{it} \mid (i,t) \in F^0]$	0.287 (0.0005)	0.213	23	$\mathbb{E}[\iota^d \mid \iota^o = 6]$	6.360 (0.0137)	6.699	43	$\vartheta_7^{EE}(6)$	0.050 (0.0011)	0.041
7	$\mathbb{E}[w_{it} \mid (i,t) \in \widehat{F}^0]$	5.192 (0.0095)	5.112	24	$\mathbb{E}[\iota^d \mid \iota^o = 7]$	6.496 (0.0156)	6.687	44	$\vartheta_7^{EE}(7)$	0.048 (0.0011)	0.037
8	sd$[w_{it} \mid (i,t) \in \widehat{F}^0]$	0.308 (0.0083)	0.218	25	$\mathbb{E}[\iota^d \mid \iota^o = 8]$	6.568 (0.0153)	6.688	45	$\vartheta_7^{EE}(8)$	0.037 (0.0009)	0.035
9	$\mathbb{E}[\Delta w_{it} \mid j(i,t) = j(i,t-1)]$	0.011 (0.0001)	0.013	26	$\mathbb{E}[\iota^d \mid \iota^o = 9]$	6.770 (0.0148)	6.631	46	$\vartheta_7^{EE}(9)$	0.037 (0.0009)	0.032
10	$\overline{w}/\underline{w}$	1.854 (0.0011)	1.800	27	$\mathbb{E}[\iota^d \mid \iota^o = 10]$	7.915 (0.0148)	6.600	47	$\vartheta_7^{EE}(10)$	0.033 (0.0009)	0.030
11	$\mathbb{E}[d_{i\tau}^{EU} = 1 \mid \tau \geq 0]$	0.338 (0.0008)	0.396	28	$\vartheta_1^{EE}(1)$	0.155 (0.0013)	0.130	48	$\vartheta_1^{EU}(1)$	0.107 (0.0011)	0.086
12	$\mathbb{E}[\overline{\tau}_i^0 \mid (i,t) \in \widehat{F}^0]$	1.054 (0.0405)	1.563	29	$\vartheta_1^{EE}(2)$	0.145 (0.0012)	0.096	49	$\vartheta_1^{EU}(2)$	0.079 (0.0010)	0.066
13	sd$[\overline{\tau}_i^0 \mid (i,t) \in \widehat{F}^0]$	1.263 (0.0829)	1.396	30	$\vartheta_1^{EE}(3)$	0.132 (0.0010)	0.081	50	$\vartheta_1^{EU}(3)$	0.063 (0.0008)	0.056
14	corr$(\overline{\tau}_i^0, w_{it} \mid (i,t) \in \widehat{F}^0)$	-0.186 (0.0316)	-0.073	31	$\vartheta_1^{EE}(4)$	0.120 (0.0011)	0.072	51	$\vartheta_1^{EU}(4)$	0.053 (0.0008)	0.050
15	Avg. jobs per emp. spell	2.192 (0.0022)	2.299	32	$\vartheta_1^{EE}(5)$	0.123 (0.0012)	0.059	52	$\vartheta_1^{EU}(5)$	0.050 (0.0008)	0.041
16	Std. dev. job per emp. spell	1.551 (0.0036)	1.439	33	$\vartheta_1^{EE}(6)$	0.114 (0.0011)	0.051	53	$\vartheta_1^{EU}(6)$	0.041 (0.0007)	0.035
17	Unemployment rate	0.218 (0.0004)	0.239	34	$\vartheta_1^{EE}(7)$	0.108 (0.0012)	0.045	54	$\vartheta_1^{EU}(7)$	0.037 (0.0007)	0.029
				35	$\vartheta_1^{EE}(8)$	0.093 (0.0010)	0.042	55	$\vartheta_1^{EU}(8)$	0.032 (0.0006)	0.025
				36	$\vartheta_1^{EE}(9)$	0.077 (0.0009)	0.037	56	$\vartheta_1^{EU}(9)$	0.024 (0.0005)	0.023
				37	$\vartheta_1^{EE}(10)$	0.052 (0.0007)	0.033	57	$\vartheta_1^{EU}(10)$	0.014 (0.0004)	0.020

Note: Standard errors in parentheses (computed by bootstrap). Hazard rates (moments 28-57) are quarterly.

the sorting related statistics do indeed vary as expected with the degree of sorting in the model.

References

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67, no. 2: 251–334.

Bagger, J., B. J. Christensen, and D. T. Mortensen (2014). Productivity and wage dispersion: Heterogeneity or misallocation. *Working paper*.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin (2014). Tenure, experience, human capital and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review* 104, no. 6: 1551–96.

Barlevy, Gadi and H. N. Nagaraja (2013). Estimating mobility rates in search models with initial condition problems. *Working paper*.

Bartolucci, Christian and Francesco Devicienti (2015). Identifying sorting in practice. *Working paper*.

Becker, Gary S. (1973). A theory of marriage: Part I. *The Journal of Political Economy* 81, no. 4: 813–46.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2016). A distributional framework for matched employer employee data. *Working paper*.

Burdett, Kenneth (1978). A theory of employee job search and quit rates. *American Economic Review* 68, no. 1: 212–220.

Burdett, Kenneth and Kenneth L. Judd (1983). Equilibrium price dispersion. *Econometrica* 51, no. 4: 955–69.

Burdett, Kenneth and Dale T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review* 39, no. 2: 257–273.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica* 74, no. 2: 323–64.

Christensen, Bent Jesper, Rasmus Lentz, Dale T. Mortensen, George Neumann, and Axel Werwatz (2005). On the job search and the wage distribution. *Journal of Labor Economics* 23, no. 1: 31–58.

Dey, Matthew S. and Christopher J. Flinn (2005). An equilibrium model of health insurance provision and wage determination. *Econometrica* 73, no. 2: 571–627.

Eeckhout, Jan and Philipp Kircher (2011). Identifying sorting - in theory. *Review of Economic Studies* 78, no. 3: 872–906.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa (2016). Job search behavior among the employed and non-employed. *Working paper*.

Fallick, Bruce C. and Charles A. Fleischman (2001). The importance of employer-to-employer flows in the u.s. labor market. *Federal Reserve Board Finance and Economics Discussion Paper 2001-18*.

Gautier, Pieter A. and Coen N. Teulings (2006). How large are search frictions. *Journal of the European Economic Association* 4, no. 6: 1193–1225.

Gautier, Pieter A. and Coen N. Teulings (2012). Sorting and the output loss due to search frictions. *Working paper*.

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993). Indirect inference. *Journal of Applied Econometrics* 8, no. 0: S85–S118.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii (2016). Identifying sorting. *Forthcoming in Econometrica*.

Haltiwanger, John C., Henry Hyatt, and Erika McEntarfer (2016). Who moves up the firm productivity job ladder? *Working paper*.

Hornstein, Andreas, Per Krusell, and Giovanni L. Violante (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review* 101, no. 4: 2873–2898.

Hosios, Arthur J. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies* 57, no. 2: 279–98.

Jarosch, Gregor (2016). Searching for job security and the cost of job loss. Stanford University, Working Paper.

Jolivet, Grégory, Fabien Postel-Vinay, and Jean-Marc Robin (2006). The empirical content of the job search model: Labor mobility and wage distributions in europe and the US. *European Economic Review* 50, no. 4: 877–907.

Lentz, Rasmus (2010). Sorting by search intensity. *Journal of Economic Theory* 145, no. 4: 1436–1452.

Lentz, Rasmus (2014). Optimal employment contracts with hidden search. *NBER Working Papers*, no. 19988.

Lentz, Rasmus, Suphanit Piyapromdee, and Jean-Marc Robin (2016). The anatomy of the wage distribution. *Working paper*.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin (2016). Matching, sorting, and wages. *Review of Economic Dynamics* 19, no. 1: 63–87.

Lopes de Melo, Rafael (2016). Sorting in the labor market: Theory and measurement. *Forthcoming in Journal of Political Economy*.

Mueller, Andreas (2010). On-the-job search and wage dispersion: New evidence from time use data. *Economics Letters 109*: 124–127.

Nagypál, Éva (2005). On the extent of job-to-job transitions. *Working paper*.

Petrongolo, Barbara and Christopher A Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature* 39, no. 2: 390–431.

Postel-Vinay, Fabien and Jean-Marc Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70, no. 6: 2295–2350.

Rubinstein, Ariel (1982). Perfect equilibrium in a bargaining model. *Econometrica* 50, no. 1: 97–109.

Shimer, Robert and Lones Smith (2000). Assortative matching and search. *Econometrica* 68, no. 2: 343–369.

Sorkin, Isaac (2015). Ranking firms using revealed preference. *Working Paper*.

Taber, Christopher and Rune Vejlin (2016). Estimation of a roy/seach/compensting differential model of the labor market. *NBER Working Papers*, no. 22439.