

Learning Your Comparative Advantages

THEODORE PAPAGEORGIOU

Penn State University, Department of Economics

First version received July 2010; final version accepted October 2013 (Eds.)

While employed, workers learn their comparative advantage and eventually choose occupations that best match their abilities. This learning process is consistent with a number of key facts about occupational mobility, such as the offsetting worker flows across occupations, the non-random patterns of occupational transitions, and the decline of occupational switching with age. We illustrate how search frictions delay learning and lead to mismatch, thereby reducing worker productivity. Moreover, we explore how different workers perform in different occupations. Are the best workers in one occupation also the best workers in another occupation (one-dimensional model of ability)? Or are some workers good at one occupation and other workers good at a different one (comparative advantage model)? The calibration favours the model of comparative advantage, as opposed to the widely used one-dimensional ability model. We use the calibrated model to investigate how the level of unemployment benefits affects worker productivity.

Key words: Occupational mobility, Learning, Unemployment.

JEL Codes: E24, E25, J24, J62

1. INTRODUCTION

Every year, a significant percentage of workers switches occupations.¹ Most of these worker flows are offsetting, so that, for every worker exiting an occupation, another worker enters it.² Moreover, occupational mobility is not random; from a given occupation, transitions to some occupations are more likely than to others.³ Finally, young workers are significantly more likely to switch occupations.⁴

This article demonstrates, both qualitatively and quantitatively, that a simple model that assigns a key role to learning about occupational comparative advantage is consistent with the key facts outlined above. We use the model to illustrate that search frictions delay learning and lead to mismatch, thereby reducing worker productivity. We also explore how different workers perform at different occupations. Are the best workers in one occupation also the best workers in another occupation (one-dimensional model of ability)? Or are some workers good at one occupation and other workers good at a different one (comparative advantage model)?

1. Thirteen percent of workers switch one-digit occupations every year (Kambourov and Manovskii, 2008).

2. Kambourov and Manovskii (2008), Jovanovic and Moffitt (1990), and Murphy and Topel (1987).

3. See the occupational transition matrices constructed in McCall (1990) and Kambourov and Manovskii (2008), which reveal noticeable differences among their off-diagonal elements. Gathmann and Schönberg (2010) also find that one's current occupation affects future occupational choices.

4. Neal (1999), Kambourov and Manovskii (2008).

In this model, each worker is endowed with a vector of productivities. Each element corresponds to the worker's productivity in one occupation, and the productivity draws can be correlated. We call this vector the worker's type. In a frictionless world, workers enter the occupation in which they are the most productive. We introduce, however, two key frictions: First, individuals do not know their type but, rather, learn it over time based on their labour market experience. For example, a sales associate might realize that he possesses good communication skills and switch to a career in advertising. Second, there are search frictions, so that individuals cannot costlessly obtain employment or switch occupations.

This setup is consistent with the above-noted key facts. To begin with, the model generates offsetting flows across occupations. As workers learn and realize that they may be more productive elsewhere, they exit an occupation. By the same token, other workers may enter that occupation. Moreover, because workers make occupational choices based on past experience, occupational choice is not random. In addition, young workers, who have little labour market experience, are the least informed about their type and are thus less likely to choose the occupation in which they perform best. Therefore, they are more likely to switch occupations, which is consistent with the data. Our setup also is consistent with the observed return mobility, whereby a fraction of switchers return to their original occupation.⁵ In our framework, workers can return to their original occupation either because they have revised upward their beliefs about the original occupation or because they have revised downward their beliefs about their new occupation.

We assess the model's ability to account for the key stylized facts of occupational choice in two complementary ways. First, in a two-type, two-occupation version of the model, we are able to fully characterize the equilibrium and list its properties. We illustrate that search frictions delay learning and lead to occupational mismatch. In addition, we underscore the different predictions of the widely used one-dimensional model of ability, whereby the best workers perform better at all occupations, versus a model of comparative advantage (Roy, 1951). Second, we provide a quantitative assessment of the model's ability to account for these facts. Even though the two-type, two-occupation model offers valuable insights owing to its tractability, it is somewhat restrictive for the purposes of empirical analysis. Therefore, we generalize our setup to allow for a continuum of types, three occupations, and human capital accumulation. This general model can only be solved numerically. We calibrate the generalized model using the 1996 panel of the Survey of Income and Program Participation (SIPP). The two-type, two-occupation model's analytical results provide guidance for moment selection.

Our results favour the model of comparative advantage as opposed to the one-dimensional model of ability, which is most often used in the literature. As illustrated in the two-type, two-occupation version of our setup, the one-dimensional ability model predicts little overlap in the support of the wage distributions of different occupations. In contrast, in the data, the occupational wage distributions overlap and are fairly similar.

We then use our calibrated model to assess the relative importance of its key features, *e.g.* learning, search frictions, on-the-job search. In particular, we simulate the model, having shut down each of the above factors sequentially. We find that preventing workers from learning leads to a significant decline in wage growth. Similarly, shutting down search frictions leads to substantial increases in occupational mobility, especially for young workers.

Finally, using the parameters from the calibrated model, we explore the aggregate implications of this type of learning. In particular, we investigate how the level of unemployment benefits affects worker productivity. Our setup illustrates a novel channel through which unemployment is costly. The rate at which workers learn about their type affects the allocation of labour across

5. 30% of workers who switch return to their one-digit occupation within a 4-year period (Kambourov and Manovskii, 2008).

occupations and thus the economy's productivity. An increase in the unemployment rate, induced by an increase in the level of unemployment benefits, results in a reduction of output per employed worker. Workers spend more time unemployed than employed and learning about their abilities. As a result, they are, on average, more likely to be mismatched. We also calculate the welfare-maximizing level of unemployment benefits and find that it is not far from the baseline specification of 30% of the average wage.

This article contributes to a growing body of literature that emphasizes the importance of occupational matching for worker outcomes (Kambourov and Manovskii (2009a) and (2009b); Gathmann and Schönberg (2010); Groes *et al.* (2010); Antonovics and Golan (2012)). Motivated by the evidence in Farber and Gibbons (1996) and Altonji and Pierret (2001), our driving force is workers' learning about their unobserved ability. This learning mechanism is better able to account for the offsetting worker flows across occupations than is a model that focuses on aggregate shocks.

This article builds on a large body of literature on workers' learning about their latent match quality, beginning with Jovanovic (1979), Miller (1984), and McCall (1990). The two-type, two-occupation model presented in Section 2 is closer to that of Moscarini (2005), who extends Jovanovic's model to allow for search frictions and derives closed-form solutions. In his work, the worker learns about the underlying quality of his match with a firm, which can be either high or low. The two-type, two-occupation model of the present article also has a similar binary restriction, which implies that beliefs follow a Bernoulli distribution. This allows us to derive closed-form expressions for the value functions and the distribution of beliefs. Related work by Felli and Harris (1996) examines how firm competition allows the worker to receive returns from his firm-specific human capital in an economy without search frictions.

Unlike Jovanovic (1979) and Moscarini (2005), the present study allows for worker productivities across occupations to be correlated. In our model, output realizations from the current match also are informative about the worker's match with other occupations. As a result, workers who separate from their match do not search randomly. Instead, they direct their search toward their preferred occupation. In addition, unlike the previous models, after revising their beliefs, workers may return to an occupation that they have abandoned.

In this respect, our setup follows that of Harris and Holmstrom (1982) and MacDonald (1982) as well as Altonji (2005). Unlike MacDonald, we focus on occupations, rather than on industries, and explore the impact of search frictions on worker productivity and the nature of sorting patterns. We also empirically assess this mechanism and incorporate several realistic features in the model: (a) different occupations can have different speeds of learning, that workers take into account in their decisions; (b) we include search frictions; and (c) we allow for general human capital accumulation. Eeckhout and Weng (2010) also consider a setup similar to the two-type, two-occupation model of the present study, but without search frictions.

Our calibration departs from other empirical articles in the literature (Gibbons *et al.* (2005); Groes *et al.* (2010); Antonovics and Golan (2012)) in that we do not restrict our attention to the one-dimensional model of ability. Instead, we also allow for a model of comparative advantage, as in Roy (1951).^{6,7} In Section 3.4, we discuss the different predictions of the two setups and how the results from our empirical exercise relate to those in Groes *et al.* (2010).

This article is organized as follows: The next section presents the two-type, two-occupation version of the model, characterizes the equilibrium as well as provides a discussion of model implications. Section 3 presents the quantitative exercise, which is based on the more general

6. See also Moscarini (2001).

7. Recent work by Sanders (2010) argues that uncertainty about abilities plays a significant role in explaining worker mobility across different tasks. Gorry (2011) argues that learning from past experience can account for the observed decline in the job-finding rate as workers grow older.

model, while Section 4 provides the model's policy implications, and Section 5 contains concluding remarks. The Appendix, as well as the Supplementary Appendix, available on the author's website, provide detailed derivations and proofs.

2. TWO-TYPE, TWO-OCCUPATION MODEL

We begin with the case in which there are two occupations and two worker types. We focus on this setup because it can be solved analytically and, therefore, clearly demonstrate the key economic mechanisms at work. In Section 3, we provide a general version of the model, with a continuum of types and three occupations.

We have three objectives. First, we illustrate that search frictions delay learning and lead to occupational mismatch. In Section 3, we quantify the impact of search frictions on worker productivity. Second, we explore whether sorting takes place according to a one-dimensional model of ability or according to comparative advantage. The two-type, two-occupation model sheds light on which moments can be used to distinguish between the two cases. These moments are used in Section 3 to identify and back out the relative productivity distribution parameters. Third, we show analytically that this type of learning can replicate the well-known fat Pareto tails of the within-occupation wage distributions.

The basic environment is the following: there are two occupations, blue collar and white collar. Workers are assumed to be of two types, with each type being most productive in a different occupation. We identify a worker's type by the occupation in which he has an advantage. Thus, a "white" type is more productive in a white-collar job than in a blue-collar job, and a "blue" type is more productive in a blue-collar job than in a white-collar one. In a frictionless world, white workers enter the white-collar occupation and blue workers enter the blue-collar occupation. We introduce, however, two key frictions: first, information frictions, whereby workers do not know their type, and, second, search frictions.

Below, we describe our setup in detail.

2.1. *Environment*

Time is continuous. There is a population of risk-neutral workers of mass one and a measure of firms. Workers die at a Poisson rate γ , while new workers are born at the same rate, which ensures that the total population remains constant. Both firms and workers share a common discount rate equal to r .

There are search frictions: firms and workers (both employed and unemployed) have to search for each other. Workers can be either employed or unemployed, and the flow value of leisure while unemployed is b_u . Workers are born unemployed. Workers, both unemployed and employed, can search in only one occupation and have to choose in which occupation to search. Search is costless. An unemployed worker meets firms at an exogenous rate λ (we endogenize λ in the Supplementary Appendix by assuming free firm entry in each occupation, subject to an entry cost). An employed worker in occupation i meets other firms, at rate $\eta_i \lambda$, where $\eta_i > 0$. Worker-firm matches dissolve exogenously at rate δ_i (e.g. plant closing), after which, the worker becomes unemployed. As shown in the next section, there also is endogenous dissolution of matches that occurs when learning leads the worker and the firm to update their value of the match surplus to a negative value.

A worker's type is an element in $\{w, b\}$, which is drawn at birth. We refer to w types as white and b types as blue. We assume that each worker's type is unknown. New-born workers draw their type from a Bernoulli distribution, where the probability of a white type is p_0 ; p_0 is common

knowledge. In the new worker population, p_0 is distributed according to a beta distribution with known shape parameters $\psi_1 > 0$ and $\psi_2 > 0$.

Each firm belongs to one of two occupations, W or B .⁸ The flow match output produced from a match κ , between a worker of type τ , $\tau \in \{w, b\}$, and a firm in occupation i , $i \in \{W, B\}$, is determined by

$$dY_t^\kappa = a_i^\tau dt + \sigma_i dZ_t^\kappa,$$

where Z_t^κ is a match-specific Wiener process, a_i^τ is the occupation and type specific mean, and σ_i is the occupation-specific output noise.⁹ Match output realizations are common knowledge. We assume, without loss of generality, that type white (w) workers are more productive in occupation W than in occupation B ($a_W^w > a_B^w$). Further, we assume that type blue (b) workers are more productive in occupation B than in occupation W ($a_B^b > a_W^b$). The case in which both types are most productive in the same occupation is not interesting, as then all workers choose that occupation, regardless of their beliefs.

Search frictions generate rents to realized matches, which are split between workers and firms via a wage-setting mechanism. Following the literature, we assume that wages are determined by generalized Nash bargaining, with $q \in (0, 1)$ denoting the worker's bargaining power.¹⁰

In the absence of output noise ($\sigma_i = 0$), firms and workers perfectly observe the worker's productivity in occupation i and thus learn about his type (as long as $a_i^w \neq a_i^b$). If, however, $\sigma_i > 0$, they observe productivity imperfectly. After observing flow match output, dY_t^κ , the market and the worker update their belief in regard to the worker's type, τ , using Bayes' rule. Beliefs follow a Bernoulli distribution, as there are only two possible types. The probability that the worker's type is white, p_t , constitutes a sufficient statistic for both the worker's output history and his initial belief. As in Liptser and Shyryaev (1977), using Bayes' rule and Ito's lemma, the belief process is reduced to

$$dp_t = p_t(1-p_t)\zeta_i \frac{dY_t^\kappa - (p_t a_i^w + (1-p_t)a_i^b)dt}{\sigma_i},$$

where $\zeta_i \equiv \frac{a_i^w - a_i^b}{\sigma_i}$. The evolution of beliefs depends on the realized match output outcome, dY_t^κ , relative to the expected one, $p_t a_i^w + (1-p_t)a_i^b$. Depending on whether a_i^w is greater or less than a_i^b (and thus whether ζ_i is positive or negative), a surprisingly high-output realization updates p_t to either a higher or a lower value, which means that the worker is either more or less likely to be a white type. The magnitude of this response depends on three factors: the current variance of beliefs $p_t(1-p_t)$, the informativeness of the output realization (signal-to-noise ratio, ζ_i), and the level of output noise, σ_i , separately.

Note that, as workers update their beliefs, they revise not only their expected performance in their current occupation but also their expected productivity in the other occupation (even if they have never been employed in it).

8. It is straightforward to extend the model to more than two occupations but still with two worker types.

9. One may imagine that the true underlying function that maps worker's type to output is more complicated. If that is the case, then learning becomes more difficult, and thus σ_i , in this case captures both output volatility as well as the (unmodelled) complexity of the production function.

10. We are abstracting from any sources of asymmetric information and view this as a natural benchmark. For an example of a three-period labour market model with firm asymmetric information, the interested reader should refer to Eeckhout (2006). See also the results in Schönberg (2007).

2.2. Equilibrium

We consider the stationary equilibrium of the above economy. We analyze equilibrium wages, optimal switching behaviour, and the stationary distribution of workers by expected productivity. For simplicity, we drop the time subscripts.

2.2.1. Wages. Because p summarizes the worker's and the firm's beliefs about the former's type, it captures their expectation for the future value of their match and serves as a state variable for their value functions, as well as for the bargained wage. As beliefs about the worker's type change, so does the worker's wage, as both the worker and the firm revise their value of the match's surplus.

Let $J_i(p)$ denote the asset value of the firm that employs a worker in occupation i , $V_i(p)$ denote the value of a worker employed in occupation i , and $U(p)$ denote the value of an unemployed worker, given beliefs p about the worker's type. Then the (cooperative) outcome of the negotiation between the worker and the firm is given by generalized Nash bargaining, which dictates that the worker's wage in occupation $i \in \{W, B\}$, $w_i(p)$ as a function of beliefs p

$$w_i(p) = \arg \max_w [J_i(p)]^{1-q} [V_i(p) - U(p)]^q. \quad (1)$$

If a worker meets another firm while employed, he chooses the firm where his value is the highest, when receiving the wage resulting from Nash bargaining, (1). In Appendix A, we show that this is an equilibrium of an ascending auction in which the current and poaching firms place bids to attract the worker. Moreover, a worker never switches to a different firm in the same occupation, as he is equally valuable to the current and poaching firm.

In this setup, the solution to the Nash bargaining problem results in the linear sharing rule

$$qJ_i(p) = (1-q)[V_i(p) - U(p)], \quad (2)$$

which provides the necessary condition to determine the worker's wage. In Appendix A, we also show that this linear sharing rule is bilaterally efficient.

2.2.2. Value functions. Below, we discuss the worker and firm value functions.

The process that governs the change in beliefs is a diffusion without a drift. Thus using Ito's lemma, we can write the flow value of an employed worker in occupation $i \in \{W, B\}$ as

$$\begin{aligned} rV_i(p) = & w_i(p) + \frac{1}{2} \xi_i^2 p^2 (1-p)^2 V_i''(p) \\ & - \delta_i [V_i(p) - U(p)] - \gamma V_i(p) + \max \{ \eta_i \lambda (V_k(p) - V_i(p)), 0 \}, \end{aligned} \quad (3)$$

where r denotes the worker's and the firm's discount rate, V_i is the value of the worker in occupation i , w_i is the occupation-specific wage function, and U the asset value of an unemployed worker.

While employed, the worker receives wage $w_i(p)$. At the same time, he benefits from learning about his type, which will allow him to make more informed decisions in the future. The second term, on the right-hand side, captures this value of learning, which depends on the variance of beliefs, $p(1-p)$, and the signal to noise ratio, ξ_i . In our setup, in contrast to that of Jovanovic (1979) and Moscarini (2005), the value of learning extends beyond the duration of the current occupational match: What the worker (and the market) learns about himself in occupation i

is also useful when he is unemployed or employed in another occupation, $k \neq i$. The worker exogenously loses his job and becomes unemployed at rate δ_i , in which case, his value is reduced to $U(p)$, which also depends on p . He dies at rate γ . Finally, at rate $\eta_i \lambda$, he meets a firm in the other occupation, which, if it is profitable for him to do so, he joins. Even though searching on-the-job is costless, it still may be optimal for the worker not to engage in on-the-job search (for instance, as shown in Appendix A, a worker has no incentive to search in the same occupation).

Note that it is possible that a worker accepts a wage cut when switching occupations, if he is compensated with a higher value of learning in his new occupation.¹¹ We return to this prediction of the model when discussing the results of the calibration presented in Section 3.4. In Appendix B, we similarly derive the value to the firm of a filled vacancy, $J_i(p)$, and the worker's value of being unemployed, $U(p)$.

In this economy workers make the following decisions: when unemployed, they choose in which occupation to search. When employed, they decide whether to search on-the-job for employment in the other occupation and when to optimally quit to unemployment. Below we provide a characterization of the optimal decision rules. Detailed derivations can be found in Appendix B as well as in Section 1 of the Supplementary Appendix.

2.2.3. Worker behavior and stationary distribution. As described above, while employed, the worker observes his match output and updates his beliefs. Should the output signals suggest that he is not employed in the occupation in which he is most productive, the worker begins searching for a job in the other occupation, while still employed. If output signals further reinforce this belief, the worker quits his current job, becomes unemployed, and concentrates his efforts on finding a job in the other occupation.

Proposition 1. *There exist unique cutoffs $\underline{p} \leq \hat{p} \leq \bar{p}$, such that:*

In the case of an unemployed worker:

- if $p \in [0, \hat{p}]$, he searches for employment in occupation B,
- if $p \in (\hat{p}, 1]$, he searches for employment in occupation W.

In the case of a worker employed in occupation B:

- if $p \in [0, \hat{p}]$, he doesn't search,
- if $p \in (\hat{p}, \bar{p})$, he searches on-the-job for employment in occupation W,
- he quits to unemployment when p reaches \bar{p} .

In the case of a worker employed in occupation W:

- if $p \in (\hat{p}, 1]$, he doesn't search,
- if $p \in (\underline{p}, \hat{p}]$, he searches on-the-job for employment in occupation B,
- he quits to unemployment when p reaches \underline{p} .

The cutoffs \underline{p} , \hat{p} and \bar{p} are determined by the solution of the system of equations (B.11)-(B.15) in Appendix B.

11. For evidence of job-to-job mobility that involve wage cuts, see Postel-Vinay and Robin (2002) and Lopes de Melo (2007).

In Appendix B, we also solve for the wage function as well as for the value of the firm, J_i , and derive the necessary conditions to solve for the undetermined coefficients. Both the wage function and the value of the firm differ, depending on whether the worker is searching on-the-job. The reason is that, when the worker leaves his current firm for a firm in another occupation, the separation is no longer bilaterally efficient, as there are lost rents for the incumbent firm. This generates a discontinuity in the wage function.

Finally, in Appendix C, we derive analytically the stationary distribution of workers by expected productivity.

Proposition 2. *The unique stationary distribution of posterior beliefs is given by equations (C.20) through (C.22) of Appendix C for workers employed in occupation B, while for those employed in occupation W, it is given by equations (6A) through (8A) of the Supplementary Appendix. The distribution of workers' posteriors in occupation W (B) features a fat right (left) Pareto-type tail, if and only if $\gamma > \zeta_W^2 \frac{\lambda+\gamma}{\delta_W+\lambda+\gamma}$ or approximately $\gamma > \zeta_W^2$ ($\gamma > \zeta_B^2$).*

These conditions have an intuitive economic interpretation: because beliefs are reset only upon death, if the speed of learning of an occupation is slower than the death rate, workers' beliefs are less likely to reach the extremes of their support before being forced to reset. In addition, in Appendix C, we derive conditions under which the within-occupation cross-sectional distribution of wages features a fat Pareto-type tail that we observe in the data.

2.3. Implications

In this section, we examine the sorting patterns that can arise in this setup. We then turn to the role of search frictions in creating occupational mismatch and reducing learning.

In the two-type, two-occupation framework, three cases of interest arise. In the first, each worker-type is better than the other in his preferred occupation. A white type, who is, by assumption, better in W than in B ($a_W^w > a_B^w$), is also better than the blue type in W ($a_W^w > a_W^b$). Similarly, a blue type, who is most productive in B ($a_B^b > a_W^b$), is also better than a white type in B ($a_B^b > a_B^w$). This case is depicted in the left graph of Figure 1, where we plot a worker's expected output in each occupation as a function of his belief. As the worker's posterior, p , approaches 1, his expected output increases in occupation W and decreases in occupation B and vice versa.

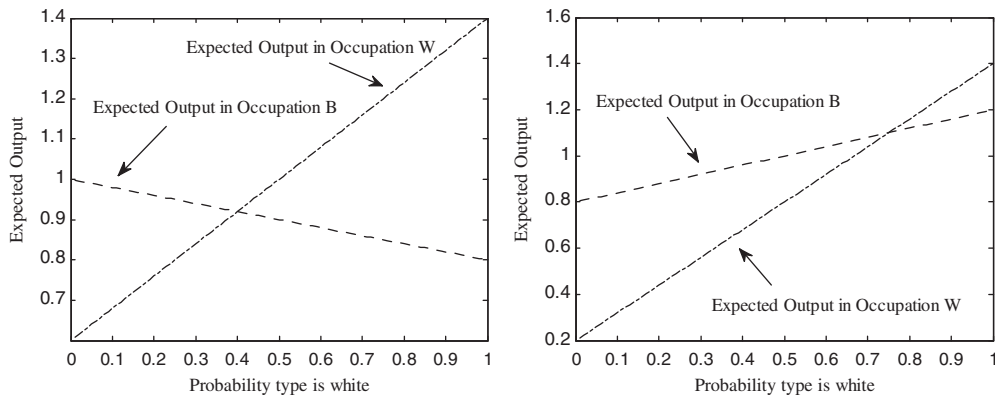


FIGURE 1

Expected output as a function of beliefs

In the second case, one type, *e.g.* the white type, has absolute advantage in both occupations, *i.e.* $a_W^w > a_W^b$ and $a_B^w > a_B^b$ (“high ability”/“low ability” case). Now, as depicted in the right graph of Figure 1, as p approaches 1, expected output in both occupations increases (although expected output in W is still larger than in B , as in the first case).

Finally, in the third case, the two types are equally productive in one of the occupations. There is no learning in that occupation, which becomes an absorbing state. In this case, as in Jovanovic (1979), the worker learns only about the quality of his match in the other occupation.

One could differentiate between the above-noted cases by looking at the support of the within-occupation wage distribution. In the case of “high ability”/“low ability” workers, our setup predicts little overlap in the wage distributions. Workers who are believed to be “high ability” work in one occupation and earn higher wages than do workers in the other occupation. This should be even more pronounced when looking at younger workers who have yet to accumulate human capital. Following this insight, in Section 3, we use the cross-sectional wage distributions to back out the relative productivity parameters.

Note also that, if search frictions are not too large, we can use information on the location in the within-occupation wage distribution of workers who switch to distinguish between the two cases. When no worker-type has an absolute advantage, workers who switch are those paid below the mean of their within-occupation wage distribution. They also are paid below the mean of the wage distribution of their new occupation. In contrast, in the “high ability”/“low ability” case, workers who switch either go from above the mean of their wage distribution to below the mean of their new occupation’s wage distribution, or they go from below the mean of their within-occupation wage distribution to above the mean of their new occupation. We examine this implication in the quantitative exercise presented in Section 4.¹²

We now turn to the role of search frictions. In this setup, search frictions lead to occupational mismatch. Consider, for instance, workers employed in occupation W with beliefs $p \in (\underline{p}, \hat{p})$. In a frictionless world, these workers would be employed in occupation B . Search frictions, however, force them to be in a suboptimal occupational match (see Proposition 1). They can only search on-the-job or quit to unemployment and search for a job in occupation B . A similar argument holds for workers employed in occupation B with beliefs $p \in (\hat{p}, \bar{p})$.

In addition, search frictions delay learning. Workers value information about their type, as illustrated by the second term in the right-hand side of the worker’s value function (equation (3)). While unemployed, however, workers do not learn about their type. As a result, they are less informed and, therefore, are mismatched more often. We revisit these productivity losses in Section 4, where we quantify the impact of increased search frictions on labour allocation.

The setup developed thus far considers the two-type, two-occupation case. We next relax this assumption in order to bring a more general model to the data.

3. QUANTITATIVE ANALYSIS

In this section we present the quantitative analysis. We assess the extent to which a relatively simple model can replicate the key features of the data. In particular, we focus on the observed occupational choices, the magnitude and the decline in occupational mobility, and the observed wage distributions. We also explore whether sorting across the occupational groups considered

12. In addition, the two cases also have different implications for wage growth. When no worker-type has an absolute advantage, wages of both worker-types increase in the long run. In contrast, in the “high ability”/“low ability” case, the “low ability” workers see their wages decline, as their type is revealed. Of course, if there is general human capital accumulation, wages of both worker-types may increase, albeit at different rates.

takes place according to a one-dimensional ability model or according to comparative advantage. Finally, we use the recovered parameters to investigate the cost of search frictions and to quantify the effect of an increase in the unemployment benefits on the economy's productivity.

Even though the two-type, two-occupation model offers valuable insights thanks to its tractability, it is somewhat restrictive for the purpose of empirical analysis. Therefore, we begin this section by generalizing our setup to allow for a continuum of types, three occupations, and human capital accumulation. Next we describe the data employed, present the calibration procedure and the results. The Section 4 discusses policy implications.

3.1. *General Model*

We extend our setup in the following dimensions: (a) we relax the binary restriction by assuming that the workers' productivity (type) is drawn from a trivariate normal distribution; (b) we allow for three instead of two occupations; and (c) we allow for general human capital accumulation. The derivations below hold for the $N > 3$ occupations case as well, but when we take our model to the data, we focus on three occupations. In this section, we discuss how these changes alter our setup.

The extended model is set in discrete rather than in continuous time. We next specify the timing of events.

First, consider a worker who is unemployed in the beginning of the period:

- a. He receives b_u this period.
- b. With probability λ , next period he is employed in occupation j , while, with probability $1 - \lambda$, he remains unemployed.
- c. With probability γ , he dies.

Consider a worker beginning the period employed in occupation i . He makes a choice between employment in occupation i , searching on-the-job (and, if failing to find another job, working in occupation i), and unemployment. If he chooses to search on-the-job:

- a. This period, he works for a firm in another occupation, j , with probability $\eta_i \lambda$, and, with probability $1 - \eta_i \lambda$, he works in his current occupation, i .
- b. He receives his wage.
- c. Production takes place, beliefs are updated, and the worker's human capital increases.
- d. With probability γ , he dies, and with probability δ_i , he becomes unemployed.

If he chooses (continuing) employment in occupation i , he receives his wage, production takes place, beliefs and human capital are updated, and the death and job loss shocks are realized. If he chooses unemployment, he simply follows the timing of an unemployed worker.

We now discuss how a worker's productivity in each occupation is determined. As in our benchmark model, we assume that workers first draw their initial belief and then draw their underlying productivity. In particular, at birth, each worker k draws his initial belief v_i^k about each occupation i

$$v_1^k \sim N(\mu_1, \kappa_1^2)$$

$$v_2^k \sim N(\mu_2, \kappa_2^2)$$

$$v_3^k \sim N(\mu_3, \kappa_3^2).$$

The realizations, v_i^k , are common knowledge. His true productivity (type), \mathbf{m}^k , is then drawn from the following distribution

$$\mathbf{m}^k = \begin{bmatrix} m_1^k \\ m_2^k \\ m_3^k \end{bmatrix} \sim N \left(\begin{bmatrix} v_1^k \\ v_2^k \\ v_3^k \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho_{12}\tau_1\tau_2 & \rho_{13}\tau_1\tau_3 \\ \rho_{12}\tau_1\tau_2 & \tau_2^2 & \rho_{23}\tau_2\tau_3 \\ \rho_{13}\tau_1\tau_3 & \rho_{23}\tau_2\tau_3 & \tau_3^2 \end{bmatrix} \right),$$

and is unobserved.

A positive correlation coefficient, ρ_{ij} , signifies that workers who have a high productivity draw in occupation i , relative to the mean, v_i^k , are likely to have a high draw relative to the mean in occupation j . In contrast, a negative correlation coefficient, ρ_{ij} , implies that a worker who is more productive than the mean in occupation i is likely to be less productive than mean in j , all else equal. In the case where $\rho_{ij} = 1$, for all i and j , the variance-covariance matrix becomes singular.¹³ In other words, there's a linear dependency among the components of \mathbf{m}^k . In this case, this implies that m_2^k and m_3^k , are linearly determined by m_1^k . This reduces to the one-dimensional model of ability, where the best workers are, on average, better at *all* occupations.

In this model, we also allow for general human capital accumulation. In particular, human capital is captured by a deterministic function, $g_{HC}(x)$, where x is working experience, *i.e.* the amount of time that a worker has spent employed.

The match output produced from a match κ between worker k and a firm in occupation i is given by

$$y_t^\kappa = m_i^k + g_{HC}(x) + \sigma_i \varepsilon_t^\kappa,$$

where $\varepsilon_t^\kappa \sim N(0, 1)$.

As in our benchmark model, a worker observes his match output and updates his beliefs about his underlying type using Bayes' rule. His beliefs in regards to his type \mathbf{m}^k also follow a trivariate normal distribution, with mean μ and variance Σ . In Section 4.1 of the Supplementary Appendix, we describe the updating of beliefs.

Before presenting the value functions, it is useful to define the value $C(\mu, \Sigma, x)$. This represents the value of a worker with beliefs (μ, Σ) and experience x , who, this period, has the option of becoming employed in any occupation

$$\begin{aligned} C(\mu, \Sigma, x) &= \max \{ w_1^{NS}(\mu, \Sigma, x) + \beta(1 - \delta_1)(1 - \gamma)E_\mu V_1(\mu, \Sigma'(1), x') + \beta\delta_1(1 - \gamma)E_\mu U(\mu, \Sigma'(1), x'), \\ &\quad w_2^{NS}(\mu, \Sigma, x) + \beta(1 - \delta_2)(1 - \gamma)E_\mu V_2(\mu, \Sigma'(2), x') + \beta\delta_2(1 - \gamma)E_\mu U(\mu, \Sigma'(2), x'), \\ &\quad w_3^{NS}(\mu, \Sigma, x) + \beta(1 - \delta_3)(1 - \gamma)E_\mu V_3(\mu, \Sigma'(3), x') + \beta\delta_3(1 - \gamma)E_\mu U(\mu, \Sigma'(3), x') \}, \end{aligned}$$

where β is the discount factor, V_i is the value of an employed worker in occupation i , and U is the value of an unemployed worker. Finally, $\Sigma'(i)$ represents the value of the updated beliefs' variance-covariance matrix. As described in Section 4.1 of the Supplementary Appendix, Σ is updated deterministically and depends on the occupation i in which the worker has been employed. Similarly, x' denotes the updated working experience. In each occupation, the worker receives wage, $w_i^{NS}(\mu, \Sigma, x)$, and if he does not experience a job loss or a death shock, he begins the following period employed in his current occupation.

13. It's straightforward to show that its determinant is equal to zero.

The value function of a worker currently employed in occupation i satisfies

$$\begin{aligned} V_i(\mu, \Sigma, x) = & \max\{I(\arg\max C(\mu, \Sigma, x) = i)C(\mu, \Sigma, x) \\ & + (1 - I(\arg\max C(\mu, \Sigma, x) = i))[\eta_i\lambda C(\mu, \Sigma, x) + \\ & (1 - \eta_i\lambda)(w_i^{OTJS}(\mu, \Sigma, x) + \beta(1 - \delta_i)(1 - \gamma)E_\mu V_i(\mu, \Sigma'(i), x')) \\ & + \beta\delta_i(1 - \gamma)E_\mu U(\mu, \Sigma'(i), x'))], \\ & U(\mu, \Sigma, x)\}. \end{aligned} \quad (4)$$

If the worker is currently employed in his preferred occupation ($\arg\max C(\mu, \Sigma, x) = i$), then his value is simply equal to $C(\mu, \Sigma, x)$ above. Otherwise, he either immediately quits to unemployment, or he searches on-the-job. In the latter case, with probability $\eta_i\lambda$, he finds a job in this preferred occupation ($\arg\max C(\mu, \Sigma, x)$), and, if not, he continues working in occupation i . Note that the wage function differs according to whether the worker is searching on the job (w_i^{NS} vs. w_i^{OTJS}).

The value of an unemployed worker satisfies

$$U(\mu, \Sigma, x) = b_u + \beta(1 - \gamma)\lambda C(\mu, \Sigma, x) + \beta(1 - \gamma)(1 - \lambda)U(\mu, \Sigma, x). \quad (5)$$

The worker receives b_u . Next period, with probability λ , he begins employment in his preferred occupation, otherwise, he spends the following period unemployed. The value of a firm, $J_i(\mu, \Sigma, x)$, is similarly defined.

The wage is set by generalized Nash bargaining, where q denotes the worker's bargaining power. As in the benchmark model, the solution to the Nash bargaining problem results in the linear sharing rule

$$qJ_i(\mu, \Sigma, x) = (1 - q)[V_i(\mu, \Sigma, x) - U(\mu, \Sigma, x)]. \quad (6)$$

In Section 4.2 of the Supplementary Appendix, we derive the wage functions w_i^{NS} and w_i^{OTJS} .

The above setup is more general than our benchmark specification, but it does not admit closed-form solutions. The solution to the worker's problem, *i.e.* which is his preferred occupation, whether to search on-the-job, and when to quit to unemployment, is derived numerically.

3.2. Data

The information necessary for our quantitative exercise includes workers' wages, occupational affiliations, and employment status as well as worker transitions between occupations and to/from unemployment. The 1996 panel of the SIPP suits our purpose well. It was designed to be a nationally representative sample of households in the civilian non-institutionalized U.S. population, for which interviews were conducted every four months for four years. For each interview, information about the worker's current occupation and wage, as well as a complete weekly employment history for the past four months was recorded. Moreover, during the SIPP's covered period, the US economy was approximately in the same phase of the business cycle (1996 through 2000).

In our setup, a worker's wage is determined by learning about his type and general human capital accumulation. Mincerian regressions show, however, that a number of worker observable characteristics, such as education, gender, and race also explain differences in wages (Abowd *et al.*, 1999) and occupational mobility (Kambourov-Manovskii, 2008). Therefore, we use only subsamples of workers that share the same observable characteristics: white males with a high

school degree. We also present results for white males with a completed college education as well as for those with some college education.

We further restrict our sample by using workers up to the age of 45 years, as we expect learning about one's unobserved aptitudes to be more important for younger workers. We also exclude workers who are in the armed forces, who are self-employed, or who have a work-preventing or limiting condition. We do not include observations for workers while they are at school or for the preceding periods. Finally, we do not use observations for which labour force data have been imputed or any observations for which the occupation may have been potentially miscoded.¹⁴

Our occupational partition for workers with a college degree and those with some college education focuses on white-collar, blue-collar, and pink-collar jobs, as in Lee and Wolpin (2006). For high school graduates, we use a different partition, as they are predominantly employed in blue-collar jobs: We focus on white-collar jobs (occupation 1); blue-collar jobs that involve precision production and repairs, such as car mechanics and carpenters (occupation 2); and blue-collar jobs that involve operators and labourers, such as various machine operators (occupation 3). We also run the calibration by grouping occupations by first index of the occupational code (first three, middle three, last three). The partition of the occupational codes into these three categories can be found in Section 5 of the Supplementary Appendix. Using a finer partition is computationally prohibitive. The 1996 panel of the SIPP used dependent interviewing, which is found to reduce occupational coding error (Hill, 1994).

We use hourly wages received at the time of the interview. We control for inflation and increases in aggregate productivity by removing the occupation-specific year effects from each wage observation. We also remove all wage observations less than one 1996 real dollar. Reported wages are top-coded at \$30, which we take into account when taking our model to the data (for instance, we use the share of wages above a certain threshold as a moment).

3.3. Calibration

We assume that the data have been generated from the steady state of our model and match theoretical moments to their empirical counterparts.

The calibration proceeds in two steps, which are performed iteratively. First, we assume that workers have learned their optimal occupation and their productivity after 20 years of working experience. We recover the true productivity distributions using the occupational choices and wages of older workers.¹⁵ Second, we retrieve the speed of learning and the information content of the initial signals using wages and occupational mobility at different points of workers' careers.

Most of the model's parameters are recovered jointly in the two-step procedure, but some are calibrated independently. In particular, a period in our model lasts for 4 months and corresponds to the interval between interview waves in our data, and the discount factor, β , is set to 0.9901, which implies an annual rate of approximately 3%. Because we do not have information on firm profits, we calibrate the value of the Nash bargaining coefficient, q , and set it equal to 0.3, as we allow for on-the-job search. This value is on the high end of the estimates in Cahuc *et al.* (2006) but is lower than the estimates of models with no on-the-job search, such as in Flinn (2006). We set the death rate parameter, γ , to 0.013333, which implies an average working life

14. According to "User Note 1 for the 1996 SIPP Cross Sectional Files", the reported occupation may be incorrect for jobs that (a) were first reported in a wave in which the worker held at least one more job and (b) were held in at least one subsequent wave.

15. Using a different approach, Lange (2006) finds that employer learning about a worker's type is relatively fast.

of approximately 25 years.¹⁶ Finally, the value of home production, b_u , is calibrated to 30% of the average wage.

We calibrate the transition parameters by using the corresponding transition moments. More specifically, the flow of workers from unemployment to employment pins down the job-finding probability λ . We calibrate this parameter following Shimer (2012) (see Section 6 of the Supplementary Appendix for details). The exogenous separation probabilities, δ_i , are set to match the flow of older workers out of employment. Consistent with older workers' having learned their type, our model implies that separation to unemployment for older workers must occur due to exogenous reasons. Because the SIPP contains weekly employment information, time aggregation is not a major concern when calculating these transition moments.

In the first step of the iterative procedure, we recover the parameters governing the true productivity distribution. More specifically, in the Supplementary Appendix, we show that the productivity parameters follow the following trivariate normal distribution

$$\begin{bmatrix} m_1^k \\ m_2^k \\ m_3^k \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \kappa_1^2 + \tau_1^2 & \rho_{12}\tau_1\tau_2 & \rho_{13}\tau_1\tau_3 \\ \rho_{12}\tau_1\tau_2 & \kappa_2^2 + \tau_2^2 & \rho_{23}\tau_2\tau_3 \\ \rho_{13}\tau_1\tau_3 & \rho_{23}\tau_2\tau_3 & \kappa_3^2 + \tau_3^2 \end{bmatrix} \right). \quad (7)$$

The value and wage of older workers, who are assumed to have learned their type, is given by equations (15A) and (16A), respectively, of the Supplementary Appendix. These workers have made their occupational choice to maximize their value, based on their wage in each occupation as well as the differences in the exogenous separation rates, δ_i . Using the observed wage distributions and occupational shares of older workers, we can recover the mean and variance-covariance matrix parameters of the productivity distribution, equation (7), conditional on the human capital accumulation process, $g_{HC}(x)$, *i.e.*

$$\phi_1 = \left\{ \mu_1, \mu_2, \mu_3, \kappa_1^2 + \tau_1^2, \kappa_2^2 + \tau_2^2, \kappa_3^2 + \tau_3^2, \rho_{12}\tau_1\tau_2, \rho_{13}\tau_1\tau_3, \rho_{23}\tau_2\tau_3 \right\}.$$

To recover parameters ϕ_1 , we simulate our setup and match the following simulated moments to the observed ones for older workers: shares of each occupation, mean wage in each of the three occupations, variance and skewness of the wage distribution in each of the three occupations, and share of wages over \$22 in each occupation. Heckman and Honoré (1990) discuss the identification of a similar setup with two occupations, but without search frictions, using the occupational shares and wage data (see also the discussion in French and Taber, 2011). In our simulation, we top-code the wages to match the observed data. For this stage, we use observations only from the first wave.

We assume that the worker accumulates human capital during his first 20 years of labour market experience according to a piecewise linear function with a knot at 10 years of experience

$$g_{HC}(x) = \omega_1 x + \omega_2 (x - 10) I(x \in (10, 20)).$$

In the second step, we recover the remaining parameters

$$\phi_2 = \{ \kappa_1, \kappa_2, \kappa_3, \sigma_1, \sigma_2, \sigma_3, \eta_1, \eta_2, \eta_3, \omega_1, \omega_2 \},$$

i.e. the standard deviation of the initial beliefs, κ_i , the standard deviation of output realizations, σ_i , the parameters that govern the effectiveness of on-the-job search, η_i , and the parameters

16. Using a smaller value for γ would imply that the average age in the cross-section of workers is unrealistically large.

associated with the human capital accumulation process, $g_{HC}(x)$, ω_1 and ω_2 . Note that, from step 1, we have pinned down $\kappa_i^2 + \tau_i^2$ and $\rho_{ij}\tau_i\tau_j$, so that, by choosing κ_i in step 2, we also determine τ_i and ρ_{ij} .

We recover the parameters ϕ_2 , by matching the following moments with their empirical counterparts: For workers with fewer than 2 years of labour market experience we use the shares as well as the mean wages of each occupation and the variance and skewness of the wage distribution for each occupation. In addition, we include the occupational switching rates for each occupation for workers with fewer than 4 years of labour market experience as well as for workers with 4–8 years of potential experience. The remaining moments consist of the shares of switches for each occupation that are job-to-job (*i.e.* without an intervening unemployment spell longer than a week) as well as the mean wage of workers after 10 years of experience. Because the choice of the human capital parameters affects the first-stage results, the two stages are performed iteratively. Section 7 of the Supplementary Appendix contains details of the procedure.

Although a rigorous identification argument is impossible, due to the complexity of our framework, we attempt to provide an informal argument on how each second-stage parameter is identified from the data. The shares, as well as the mean, the variance, and skewness of the wages of young workers for each occupation, pin down the standard deviations of the initial beliefs, κ_i . The rate at which workers switch occupations, as well as the rate's decrease, specify the speed at which workers learn about their underlying productivities, which is governed by σ_i . The fraction of job-to-job transitions pins down the efficiency of on-the-job search, η_i . The overall level of wages for young workers, as well the mean wage after 10 years of experience, determines the rate of human capital accumulation. Wage growth depends on both learning and sorting of workers into occupations as well the rate of human capital accumulation. Conditional, however, on the speed of learning, here determined by σ_i , we can pin down the rate of human capital accumulation.

3.4. Results

The resulting parameters are reported in Table 1, while Table 2 shows the fit.

Our results suggest that, on average, high school graduates are less productive in occupation 2 than in the other two occupations, as $\mu_2 < \mu_1, \mu_3$. The standard deviation of the productivity draw, τ_2 , however, is larger for that occupation. The large standard deviation of productivity

TABLE 1
Model parameters: mean productivities, μ_i , standard deviation of productivities, τ_i , dispersion of initial beliefs, κ_i , human capital parameters, ω_1 and ω_2 , correlation coefficients, ρ_{ij} , signal noise parameters, σ_i , efficiency of on-the-job search parameters, η_i , exogenous separation probabilities, δ_i , job finding probability, λ , and value of home production, b_u

μ_1	7.08	ρ_{23}	−0.934
μ_2	6.09	σ_1	12.01
μ_3	7.01	σ_2	14.44
τ_1	4.02	σ_3	9.99
τ_2	10.35	η_1	0.35
τ_3	8.18	η_2	0.204
κ_1	3.66	η_3	0.199
κ_2	1.03	δ_1	0.017
κ_3	1.02	δ_2	0.018
ω_1	0.157	δ_3	0.016
ω_2	−0.28	λ	0.828
ρ_{12}	0.011	b_u	3.33
ρ_{13}	−0.161		

TABLE 2

Targeted moments. “Young” refers to workers between the age of 18 and 20 years, while “Old” refers to workers with more than 20 years of potential labour market experience (up to the age of 45 years). Occupational switching rates are 8-month rates. They are computed during the first and second 4-year periods of labour market experience

Moment	Data	Predicted	Moment	Data	Predicted
Old Empl Share 1	21.42%	20.99%	Young Mean Wage 3	\$7.13	\$5.35
Old Empl Share 2	37.56%	38.36%	Young St Dev Wage 1	\$1.80	\$4.21
Old Mean Wage 1	\$11.62	\$11.29	Young St Dev Wage 2	\$2.92	\$8.04
Old Mean Wage 2	\$14.38	\$13.93	Young St Dev Wage 3	\$2.11	\$7.07
Old Mean Wage 3	\$12.57	\$12.57	Young Skewness 1	0.82	−0.24
Old St Dev Wage 1	\$4.94	\$3.28	Young Skewness 2	0.71	0.57
Old St Dev Wage 2	\$5.39	\$5.40	Young Skewness 3	1.11	0.21
Old St Dev Wage 3	\$4.81	\$4.57	Occ Switching 1 (0-4 years)	11.29%	14.84%
Old Skewness 1	1.09	0.48	Occ Switching 2 (0-4 years)	13.04%	15.14%
Old Skewness 2	0.65	0.62	Occ Switching 3 (0-4 years)	13.48%	14.72%
Old Skewness 3	0.91	0.70	Occ Switching 1 (5-8 years)	7.79%	6.51%
Old % >\$22 occup 1	5.63%	0.28%	Occ Switching 2 (5-8 years)	5.53%	2.92%
Old % >\$22 occup 2	7.63%	9.16%	Occ Switching 3 (5-8 years)	8.34%	2.58%
Old % >\$22 occup 3	4.04%	3.66%	% Job-to-Job Switches occup 1	69.88%	92.08%
Young Empl Share 1	28.99%	20.51%	% Job-to-Job Switches occup 2	71.43%	49.98%
Young Empl Share 2	26.82%	34.19%	% Job-to-Job Switches occup 3	73.13%	58.52%
Young Mean Wage 1	\$6.06	\$8.16	Mean Wage 30	\$10.87	\$12.06
Young Mean Wage 2	\$7.99	\$4.85			

explains the large fraction of workers who eventually choose occupation 2 (37.56% as shown in Table 2). The low dispersion of initial beliefs (κ_i) leads to a lower share in occupation 2 (26.82%) among younger workers. These workers initially believe that they may be better suited in one of the other two occupations, which have a higher mean (μ_i). The differences in the level of noise (σ_i) and the ability to search on-the-job (η_i) are small across occupations. The standard deviation of the posterior belief, which we denote $(\tau_i^k)^*$, of a worker employed in occupation 1 for 3 years falls from \$3.81 to \$2.84; for a worker employed in occupation 2, it falls from \$8.41 to \$4.37; and for a worker employed in occupation 3, it falls from \$6.33 to \$3.09.¹⁷

The human capital profile is initially upward sloping and then downward sloping, potentially indicating human capital decumulation. Bagger *et al.* (2013) also report a gradual loss of productivity.

The results also suggest that the allocation of workers across these occupations is not well described by the one-dimensional model of ability, in which some workers perform better at all occupations. As discussed in Section 3.1, a model where the best workers are better at *all* occupations implies that ρ_{ij} is equal to one for all i and j . In this case, for instance, the correlation coefficient, ρ_{23} , is equal to -0.934 .

Our findings are not surprising. As illustrated in the two-type, two-occupation model of Section 2, the one-dimensional ability model implies that there is little overlap in the occupational wage distributions. In the case of two occupations, workers who are thought to be high ability, are employed in the high-ability occupation and earn higher wages. In contrast, workers suspected of being low ability work in the low-ability occupation. In the extreme case, with no search frictions and no differences in the speed of learning, every worker in the low-ability occupation is paid less than the lowest paid worker in the high-ability occupation. In this case, there is no overlap in the support of the two wage distributions. This can be relaxed by introducing search frictions, differences in the speed of learning, occupation-specific human capital, or measurement error.

17. See updating equations in Section 4.1 of the Supplementary Appendix.

TABLE 3

Fraction of job-to-job transitions involving a wage cut. Wage growth of workers who remain in same occupation between ages of 20 and 24 years. Standard deviation of individual wage changes for workers who remain in the same occupation

	Data	Predicted
% Job-to-Job Transitions with Wage Cuts	19%	35%
Wage growth in Occupation 1	54.87%	66.74%
Wage growth in Occupation 2	34.25%	32.37%
Wage growth in Occupation 3	46.69%	13.22%
Wage Change St. Deviation	\$1.97	\$1.89

Nevertheless, an extended model would still have difficulty reconciling two occupations that have similar wage distributions with a one-dimensional model.

Indeed, the wage distributions of the three occupations do not differ much (see Table 2), and, therefore, our results do not favour the one-dimensional model of ability. The results presented here are complementary to those of Groes *et al.* (2010), who present evidence in favour of vertical sorting across occupations. Our empirical exercise focuses on three broad occupational groups, whereas Groes *et al.* look at finer classifications. Taken together, the results suggest that there are groups of occupations, within which workers sort vertically, whereas across groups, sorting takes place according to workers' comparative advantages, as shown in the above results.

Moreover, as discussed in Section 2.3 of the two-type, two-occupation model, we can explore the wages of workers who switch occupations relative to other workers in the same occupation. In the data, the average pre-switch wage of workers is lower than the mean in their corresponding occupation. Moreover, workers' post-switch wages are lower than the prevailing mean in their new occupation. This holds in both the raw data as well as after controlling for a fourth-order polynomial in age (with the exception of workers who have moved to occupation 1, who earn approximately the same as existing workers). The calibrated model also is consistent with both facts in all three occupations, although the differences between switchers and stayers are somewhat larger.¹⁸ As discussed in Section 2.3, this also is consistent with the calibration's result that the one-dimensional model of ability, in which some workers perform better at all occupations, does not describe well the allocation of workers across these occupations.

Further, as shown in Table 3, there is a significant number of switches that involve a wage cut (35%) in the calibrated model, somewhat more than in the data (19%). As shown in the two-type, two-occupation model, a worker is willing to accept a lower wage in exchange for a higher value of learning in his new occupation (see the discussion on equation (3)).¹⁹ Table 3 also presents the wage growth of workers who remain in the same occupation between the ages of 20 and 24 years, both in the model and in the data. It also shows the standard deviation of wage changes, from one period to the next, for workers who remain in the same occupation. The predicted

18. When controlling for a fourth-order polynomial in age, in the data, the residual pre-switch mean wage of switchers in occupation 1 is \$-1.16 versus \$-0.90 for stayers, \$0.08 versus \$1.26 for occupation 2, and \$-1.30 versus \$-0.25 for occupation 3. In the simulations, the corresponding residuals are \$-3.60 versus \$-0.05 for occupation 1, \$-4.38 versus \$0.90 for occupation 2, and \$-3.94 versus \$0.32 for occupation 3. Further, in the data, the residual post-switch mean wage of switchers in occupation 1 is \$-0.97 versus \$-1.01 for stayers, \$0.07 versus \$1.18 for occupation 2, and \$-1.10 versus \$-0.35 for occupation 3. In the simulations, the corresponding residuals are \$-1.20 versus \$-0.18 for occupation 1, \$-2.09 versus \$0.99 for occupation 2, and \$-1.49 versus \$0.39 for occupation 3.

19. The wage cuts in the calibration reflect the option value of learning one's ability in another occupation as well as the discrete time specification of the setup: a worker who has a low output realization is more likely to start looking for a job in another occupation. If he is immediately successful, then the low realization does not show up as a low wage in his original occupation in the following period.

TABLE 4

Model parameters: mean productivities, μ_i , standard deviation of productivities, τ_i , dispersion of initial beliefs, κ_i , human capital parameters, ω_1 and ω_2 , correlation coefficients, ρ_{ij} , signal noise parameters, σ_i , efficiency of on-the-job search parameters, η_i , exogenous separation probabilities, δ_i , job finding probability, λ , and value of home production, b_u

College				Some College				High School Alt Classification			
μ_1	11.20	ρ_{23}	-0.955	μ_1	2.95	ρ_{23}	-0.81	μ_1	1.89	ρ_{23}	-0.14
μ_2	5.73	σ_1	63.35	μ_2	14.40	σ_1	30.23	μ_2	6.58	σ_1	19.62
μ_3	7.12	σ_2	50.33	μ_3	0.07	σ_2	20.74	μ_3	8.11	σ_2	40.01
τ_1	13.18	σ_3	56.63	τ_1	10.20	σ_3	29.73	τ_1	4.56	σ_3	35.37
τ_2	8.03	η_1	0.395	τ_2	7.32	η_1	0.40	τ_2	6.82	η_1	0.32
τ_3	7.75	η_2	0.396	τ_3	9.83	η_2	0.50	τ_3	5.66	η_2	0.31
κ_1	5.63	η_3	0.423	κ_1	3.01	η_3	0.30	κ_1	3.00	η_3	0.30
κ_2	2.91	δ_1	0.005	κ_2	2.94	δ_1	0.008	κ_2	3.12	δ_1	0.015
κ_3	2.10	δ_2	0.006	κ_3	1.99	δ_2	0.008	κ_3	1.97	δ_2	0.02
ω_1	0.19	δ_3	0.005	ω_1	0.21	δ_3	0.013	ω_1	0.15	δ_3	0.015
ω_2	-0.307	λ	1	ω_2	-0.41	λ	1	ω_2	-0.2	λ	0.828
ρ_{12}	0.002	b_u	4.15	ρ_{12}	0.10	b_u	3.72	ρ_{12}	-0.10	b_u	3.32
ρ_{13}	0.275			ρ_{13}	-0.56			ρ_{13}	0.26		

standard deviation is fairly close to its empirical counterpart. Finally, it is worth noting that the setup generates significant return mobility, as 31% of switchers eventually return to their original occupation.

We repeat this exercise, using a different occupational partition. In particular, we group occupations by the first index of the occupational code (first three, middle three, last three). In addition, we present results when focusing on the subsample of workers with a college degree and those who have some college education.

The resulting parameters are presented in Table 4, while the moments matched are seen in Tables 5 through 7. Workers with a college degree are, on average, significantly more productive in occupation 1 than in the other two occupations. Nonetheless, there is still substantial dispersion (τ_i) in productivity realizations in both occupations 2 and 3. This leads to non-negligible employment shares for these occupations.

Unlike workers with a college degree, workers with some college education are, on average, more productive in occupation 2 than in occupation 1. Again, however, the large dispersion of realized productivity (τ_i) leads to high shares for the other two occupations.

When reclassifying occupations for high school graduates, we find that they are significantly less productive in the (reclassified) occupation 1. The relatively low dispersion of realized productivity in that occupation (τ_1) accounts for the corresponding low employment share.

For all three groups, as before, the human capital profile is initially upward sloping and then downward sloping.

It is also worth noting that, in all subsamples, the results suggest that the allocation of workers across these occupations is not well described by the one-dimensional model of ability. For instance, ρ_{23} is equal to -0.955 in the college subsample and -0.81 in the some college subsample, while, in the alternative classification for high school graduates, all three correlations are not far from zero.

To better understand the impact of the various elements of the model, we perform a number of counterfactual exercises, using the original calibration. The results are reported in Table 8.

First, we shut down learning and assume that workers keep their prior belief forever. Because there is no learning, workers remain in the same occupation forever. This experiment sheds light on the importance of information frictions. Wage dispersion now arises only because of

TABLE 5

Targeted moments - College. "Young" refers to workers between the age of 22 and 24 years, while "Old" refers to workers with more than 20 years of potential labour market experience (up to the age of 45 years). Occupational switching rates are 8-month rates. They are computed during the first and second 4-year periods of labour market experience

Moment	Data	Predicted	Moment	Data	Predicted
Old Empl Share 1	50.55%	46.04%	Young Mean Wage 3	\$6.41	\$8.44
Old Empl Share 2	23.08%	26.64%	Young St Dev Wage 1	\$5.12	\$4.00
Old Mean Wage 1	\$18.58	\$20.62	Young St Dev Wage 2	\$1.95	\$2.75
Old Mean Wage 2	\$14.18	\$13.38	Young St Dev Wage 3	\$2.20	\$2.38
Old Mean Wage 3	\$13.14	\$13.14	Young Skewness 1	0.81	0.28
Old St Dev Wage 1	\$7.37	\$6.20	Young Skewness 2	2.07	0.39
Old St Dev Wage 2	\$4.90	\$5.28	Young Skewness 3	-0.01	0.57
Old St Dev Wage 3	\$6.11	\$5.26	Occ Switching 1 (0-4 years)	8.07%	8.42%
Old Skewness 1	-0.10	-0.34	Occ Switching 2 (0-4 years)	38.46%	21.85%
Old Skewness 2	0.31	0.53	Occ Switching 3 (0-4 years)	9.52%	12.62%
Old Skewness 3	0.89	0.53	Occ Switching 1 (5-8 years)	3.53%	5.53%
Old % >\$22 occup 1	39.13%	45.48%	Occ Switching 2 (5-8 years)	7.14%	11.03%
Old % >\$22 occup 2	4.76%	7.00%	Occ Switching 3 (5-8 years)	3.57%	9.20%
Old % >\$22 occup 3	8.33%	6.42%	% Job-to-Job Switches occup 1	82.35%	86.50%
Young Empl Share 1	47.02%	48.76%	% Job-to-Job Switches occup 2	76.19%	83.93%
Young Empl Share 2	17.22%	21.86%	% Job-to-Job Switches occup 3	82.35%	81.22%
Young Mean Wage 1	\$10.11	\$12.30	Mean Wage 30	\$14.41	\$13.09
Young Mean Wage 2	\$6.39	\$8.39			

TABLE 6

Targeted moments - Some College. "Young" refers to workers between the age of 21 and 23 years, while "Old" refers to workers with more than 20 years of potential labour market experience (up to the age of 45 years). Occupational switching rates are 8-month rates. They are computed during the first and second 4-year periods of labour market experience

Moment	Data	Predicted	Moment	Data	Predicted
Old Empl Share 1	14.02%	12.94%	Young Mean Wage 3	\$6.40	\$6.88
Old Empl Share 2	65.61%	69.05%	Young St Dev Wage 1	\$2.77	\$5.65
Old Mean Wage 1	\$14.90	\$14.49	Young St Dev Wage 2	\$3.94	\$5.84
Old Mean Wage 2	\$15.08	\$14.72	Young St Dev Wage 3	\$2.10	\$3.91
Old Mean Wage 3	\$11.40	\$11.34	Young Skewness 1	1.19	0.07
Old St Dev Wage 1	\$5.52	\$5.91	Young Skewness 2	0.82	-0.07
Old St Dev Wage 2	\$5.61	\$4.79	Young Skewness 3	0.43	0.31
Old St Dev Wage 3	\$5.20	\$5.71	Occ Switching 1 (0-4 years)	20.83%	20.04%
Old Skewness 1	0.68	0.42	Occ Switching 2 (0-4 years)	4.44%	22.96%
Old Skewness 2	0.44	0.34	Occ Switching 3 (0-4 years)	16.67%	28.03%
Old Skewness 3	0.38	0.74	Occ Switching 1 (5-8 years)	16.67%	5.57%
Old % >\$22 occup 1	7.55%	12.28%	Occ Switching 2 (5-8 years)	3.28%	4.00%
Old % >\$22 occup 2	10.08%	11.90%	Occ Switching 3 (5-8 years)	6.41%	12.35%
Old % >\$22 occup 3	1.30%	3.05%	% Job-to-Job Switches occup 1	75.76%	77.88%
Young Empl Share 1	9.96%	33.53%	% Job-to-Job Switches occup 2	86.84%	97.80%
Young Empl Share 2	58.24%	62.94%	% Job-to-Job Switches occup 3	65.85%	75.38%
Young Mean Wage 1	\$8.44	\$8.41	Mean Wage 30	\$12.00	\$13.43
Young Mean Wage 2	\$9.37	\$11.86			

the differences in the priors and human capital accumulation. As shown in the second column of Table 8, the share of occupation 2 is reduced to 2%. The high value of τ_2 , compared to κ_2 , implies that workers who had high draws in occupation 2 do not know this and choose not to work there, given the low dispersion of priors. In our setup, learning leads to wage growth because (a) it allows workers to self-select into the occupation that they perform best; and (b) it revises upward their expected output in that occupation. When we shut down learning, wage growth is reduced significantly to 0.61% annually, as it is now driven exclusively by general human capital

TABLE 7

Targeted moments - High School-Alternative Occupational Classification. "Young" refers to workers between the age of 18 and 20 years, while "Old" refers to workers with more than 20 years of potential labour market experience (up to the age of 45 years). Occupational switching rates are 8-month rates. They are computed during the first and second 4-year periods of labour market experience

Moment	Data	Predicted	Moment	Data	Predicted
Old Empl Share 1	9.05%	9.40%	Young Mean Wage 3	\$7.18	\$7.50
Old Empl Share 2	40.88%	40.42%	Young St Dev Wage 1	\$1.48	\$1.10
Old Mean Wage 1	\$11.84	\$10.30	Young St Dev Wage 2	\$2.87	\$1.93
Old Mean Wage 2	\$13.55	\$13.64	Young St Dev Wage 3	\$2.07	\$1.74
Old Mean Wage 3	\$12.86	\$12.89	Young Skewness 1	0.84	-0.27
Old St Dev Wage 1	\$5.16	\$3.76	Young Skewness 2	0.94	0.32
Old St Dev Wage 2	\$5.58	\$4.82	Young Skewness 3	1.02	-0.10
Old St Dev Wage 3	\$4.78	\$4.34	Occ Switching 1 (0-4 years)	20.46%	12.59%
Old Skewness 1	1.12	0.10	Occ Switching 2 (0-4 years)	12.50%	10.76%
Old Skewness 2	0.73	0.26	Occ Switching 3 (0-4 years)	9.76%	5.83%
Old Skewness 3	0.84	0.18	Occ Switching 1 (5-8 years)	11.54%	11.08%
Old % >\$22 occup 1	8.34%	0.18%	Occ Switching 2 (5-8 years)	4.78%	6.12%
Old % >\$22 occup 2	7.01%	5.02%	Occ Switching 3 (5-8 years)	8.75%	4.15%
Old % >\$22 occup 3	4.22%	2.27%	% Job-to-Job Switches occup 1	68.09%	80.17%
Young Empl Share 1	11.32%	2.58%	% Job-to-Job Switches occup 2	65.57%	69.76%
Young Empl Share 2	38.14%	37.40%	% Job-to-Job Switches occup 3	75.00%	77.79%
Young Mean Wage 1	\$6.28	\$5.79	Mean Wage 30	\$10.85	\$9.24
Young Mean Wage 2	\$7.09	\$7.66			

TABLE 8
Counterfactuals

Moment	Baseline	No Learning	No Search Frictions	No OTJS	No Initial Disp
Young % occup 1	20.51%	38.14%	50.75%	31.52%	13.65%
Young % occup 2	34.19%	2.01%	33.02%	46.54%	19.77%
Young Mean Wage occ 1	\$8.16	\$9.06	\$6.67	\$7.20	\$10.96
Young Mean Wage occ 2	\$4.85	\$6.45	\$9.99	\$3.22	\$6.97
Young Mean Wage occ 3	\$5.35	\$6.57	\$12.78	\$6.57	\$5.45
Young St Dev Wage 1	\$4.21	\$0.95	\$3.37	\$3.76	\$2.94
Young St Dev Wage 2	\$8.04	\$1.60	\$4.98	\$9.89	\$6.82
Young St Dev Wage 3	\$7.07	\$0.58	\$4.30	\$7.34	\$6.67
Young Skewness 1	-0.24	0.48	-0.08	-0.02	-2.05
Young Skewness 2	0.57	1.03	1.32	0.68	0.21
Young Skewness 3	0.21	2.79	0.61	0.09	0.10
Occ Sw 1 (0-4 years)	14.84%	0	50.72%	4.79%	9.22%
Occ Sw 2 (0-4 years)	15.14%	0	31.38%	14.92%	10.33%
Occ Sw 3 (0-4 years)	14.72%	0	18.01%	7.14%	12.88%
Occ Sw 1 (5-8 years)	6.51%	0	11.85%	1.44%	5.88%
Occ Sw 2 (5-8 years)	2.92%	0	7.10%	1.16%	3.80%
Occ Sw 3 (5-8 years)	2.58%	0	6.26%	1.27%	2.37%
% Job-to-Job Sw occup 1	92.08%	NaN	1	0	77.79%
% Job-to-Job Sw occup 2	49.98%	NaN	1	0	76.33%
% Job-to-Job Sw occup 3	58.52%	NaN	1	0	56.32%
Mean Wage 30	\$12.06	\$8.72	\$18.29	\$11.97	\$11.73
Wage Growth up to 35	4.55%	0.61%	4.24%	5.16%	3.57%

accumulation. Note, however, that initial wages are higher, as workers now do not accept a lower wage in exchange for the option of learning about their type.

We then shut down search frictions and allow workers to contact potential employers at no cost. Assuming perfect competition among a large mass of potential employers, workers are now paid their expected output, while firm profits drop to zero. In this case, as shown in the third column of Table 8, both wages and occupational mobility increase.

If we shut down the ability of workers to search while employed, occupational mobility falls substantially (fourth column of Table 8). A large fraction of occupational switches involve employed workers. Imposing the extra cost of having to quit to unemployment in order to switch occupations reduces significantly the number of realized occupational transitions.

Finally, we eliminate the initial dispersion of beliefs, *i.e.* set $\kappa_i = 0$, for all i in the last column of Table 8. Now, all workers start with the same prior and thus make the same occupational choice. In particular, they prefer to start off in occupation 3, which has a high mean, μ_3 , as well as a high dispersion of productivities, τ_3 , and low noise, σ_3 . Moreover, the variance of wages during the first two years also falls. We should note that this counterfactual changes not only the information set of new workers but also alters the true productivity distribution, unlike the previous counterfactuals.

4. POLICY IMPLICATIONS

We then use our model to investigate the impact of increased unemployment on labour allocation. In the 1970s, the U.S. and several European countries experienced large increases in their unemployment rates that persisted for many years. For many countries, the unemployment rate shot up from less than 3% in the early 1970s to 10% by 1980. An increase in the percentage of workers not employed results in output loss simply because fewer resources are being utilized. We argue, however, that there is an additional cost in terms of output per employed worker.

In particular, workers now spend more time unemployed and less time employed and learning about their ability. As illustrated in the two-type, two-occupation model of Section 2, reducing the ability of workers to switch occupations implies that they are now more likely to be employed in an occupation that does not match their abilities.

To explore the impact of increased search frictions on productivity, we use the parameter values generated by the model's calibration and perform a counterfactual exercise. The thought experiment involves a simulation of an increase in the unemployment rate similar to the one that occurred in the U.S. and Europe in the 1970s. More specifically, we permanently decrease the job-finding rate, λ , to generate an unemployment rate of approximately 10%. In other words, we are comparing two steady states, which is more relevant when considering the long-run labour market differences between Europe and the US.

Table 9 presents the results of our exercise. Our baseline estimates, in column 1, imply an unemployment rate of approximately 5.8% and hourly output per employed worker of \$13.81. When we generate an increase in the unemployment rate to approximately 10%, output per employed worker drops to \$13.02 per hour.

The above result has important implications for policy makers. Economic policies that affect search frictions influence the economy's allocation of resources and thus productivity. For instance, consider a change in the flow value of leisure while unemployed, b_u , which may be due to an increase in unemployment benefits. More specifically, assuming a fixed job-finding rate, an increase in unemployment benefits raises the outside value of the worker. As a consequence,

TABLE 9
Impact of search frictions on productivity

	$b_u = 3.33$	$b_u = 3.33$	$b_u = 5.55$
	$\lambda_1 = \lambda_2 = \lambda_3 = 0.83$	$\lambda_1 = \lambda_2 = \lambda_3 = 0.44$	$\lambda_1 = 0.7, \lambda_2 = 0.75, \lambda_3 = 0.71$
Unemployment	5.83%	10.07%	6.89%
Output per Worker	\$13.81	\$13.02	\$13.55

matches that were acceptable before now yield a negative surplus, resulting in a higher separation rate. In addition, an increase in b_u decreases profits of filled vacancies and dampens incentives to post vacancies in the first place, thus leading to a lower job-finding rate.

We then explore the impact of the increase of unemployment benefits on productivity. We use the parameter values generated by the model's calibration and permanently increase the value of b_u to \$5.55, which corresponds to 50% of the average wage, as opposed to 30%. The increase in b_u alters incentives for firm entry. To endogenize vacancy posting we follow den Haan *et al.* (2000) and assume that the matching function is of the form $m(p_i, v_i) = \frac{p_i v_i}{p_i + v_i}$, where p_i is the effective mass of workers (employed or unemployed) petitioning for a job in occupation i and v_i is the mass of occupation i vacancies. Using the baseline results, we recover the costs of posting a vacancy in each occupation. When we change b_u , we recover the job-finding rates that imply zero expected profits of posting a vacancy in each of the three occupations.

As shown in the third column of Table 9, the increase of b_u leads to an increase of the unemployment rate to approximately 6.9%, and output per worker falls to \$13.55.²⁰ Moreover, we also calculate welfare in the two cases, assuming that unemployment benefits are financed through lumpsum taxes on all workers. Welfare is calculated as in Flinn (2006). It is given by the weighed sum of the average welfare of employed workers, unemployed workers, workers out of the labour force, and filled vacancies. The weights are given by the size of each group. We find that welfare is lower in the case of the increased b_u . In fact, the welfare-maximizing value of b_u is \$2.89, close to our baseline specification of \$3.33.

The results in our setup contrast with those of the literature. Both Acemoglu and Shimer (1999) and Marimon and Zilibotti (1999) have illustrated mechanisms under which such an increase leads to improved resource allocation. In Acemoglu and Shimer, increased unemployment insurance encourages risk-averse workers to apply for higher quality jobs, which are more difficult to obtain. Marimon and Zilibotti show that increased benefits act as a search subsidy that allows workers to obtain more suitable jobs. Here, the increase in unemployment insurance leads to a higher unemployment rate and to a deterioration of the allocation of workers across productive activities. Estimating a model that incorporates all the above mechanisms is beyond the scope of this article. Our results, however, partly explain why most empirical studies have had difficulty detecting the positive impact of unemployment insurance on the match quality that had been previously presumed (van Ours and Vodopivec, 2008).

5. CONCLUSION

This article demonstrates, both qualitatively and quantitatively, that a simple model that assigns a key role to learning about occupational comparative advantage is consistent with key facts of occupational mobility. In particular, the model is consistent with the offsetting worker flows across occupations, the decline of occupational mobility with age, the experience profile of wages, and the cross-sectional within-occupation wage inequality among observationally equivalent workers.

Our setup can be viewed as an equilibrium Roy (1951) model with learning. We use the model to understand the patterns of occupational sorting. We find that sorting takes place according to comparative advantage, rather than according to a one-dimensional model of ability, where the best workers go to the best occupations. Our setup also illustrates a novel channel through which unemployment is costly. In particular, it delays learning and causes mismatch by preventing workers from being employed in their preferred occupation. The calibrated model finds that the resulting productivity losses are substantial.

20. In the case of increased b_u , the fraction of workers who choose to remain out of the labour force, which is negligible in our baseline calibration, rises to 5.3%.

APPENDIX

A. ON-THE-JOB SEARCH AND WAGES

In the main text we assume that when a worker moves to another firm, his wage is given by the Nash bargaining. We show here that this is an equilibrium of an auction between the incumbent firm and the firm the worker has contacted.

The main assumptions are the following: (a) wages are set by continuous renegotiation and (b) on-the-job search is costless for the worker and unobserved (by the firm). Both of these assumptions are fairly standard and are also made for instance in Pissarides (1994).

The structure of the auction, which closely resembles the one in Moscarini (2005), is the following. When an employed worker meets a new firm there is competition for the worker's services. The competition determines the firm (incumbent or poacher) where the worker becomes employed, as well as a lumpsum transfer from the winning firm to the worker. After the competition is over, the winning firm engages in continuous renegotiation of wages with the worker.

Firm competition is according to the following protocol:

1. Participation: first, the incumbent decides whether to pay $\varepsilon > 0$ and enter the auction (yes-incumbent-stage1 or no-incumbent-stage1: Y-I1 or N-I1). Second, the poacher observes the incumbent's choice and decides whether to pay $\varepsilon > 0$ and enter the auction (Y-P or N-P). Third, if the outcome so far is (N-I1, Y-P), then the incumbent decides whether to pay $\varepsilon > 0$ and enter the auction (Y-I2, N-I2).
2. Auction: If no firm enters the auction, then the worker remains with the incumbent firm and there is no transfer. If one firm enters the auction, then the worker becomes employed at that firm and there is no transfer. If both firms enter the auction, then the transfer is determined by an ascending bid auction and the worker becomes employed at the winning firm. In case of a tie, he remains employed at the incumbent firm.

We consider the subgame perfect Nash equilibrium of this game as $\varepsilon > 0$ goes to zero. The game is solved by backward induction.

Let $J^I(p)$ the value to the incumbent firm of employing the worker with belief p and $J^P(p)$ the corresponding value to the poaching firm. This is the maximum amount that each firm is willing to bid in the form of lump-sum transfer, for the worker. In the case where both firms enter the auction, the worker receives a transfer equal to $\min\{J^I(p), J^P(p)\}$.

There are three possibilities: $J^I(p) > J^P(p)$, $J^I(p) < J^P(p)$ and $J^I(p) = J^P(p)$. See also Figures A1, A2 and A3 for a graphic representation of the solution for each of the three cases.

In the first case ($J^I(p) > J^P(p)$): If the incumbent chooses to enter the auction in the first stage, then it is optimal for the poacher to decline entry as he would lose the ensuing auction. As a result the worker stays with the incumbent and there is no transfer. If the incumbent does not enter in the first stage and the poacher does, then it is optimal for the incumbent to pay ε and enter as he would win the auction. Again the worker stays with the incumbent with no transfer. Therefore if the incumbent has chosen not to enter in the first stage, it is optimal for the poacher to decline entry as well. The poacher foresees that if he chose to enter, the incumbent would enter in the second stage and win the auction. The incumbent, foreseeing that if declines to enter the auction in the first stage, so will the poacher, chooses to not to enter. The subgame perfect equilibrium here is (N-I1, N-P), where neither firm enters the auction, the worker stays with the incumbent with no transfer.

In the second case ($J^I(p) < J^P(p)$): If the incumbent chooses to enter the auction in the first stage, then it is optimal for the poacher to enter the auction, since he knows that he will win the auction. The worker moves to the poacher and receives a transfer equal to $J^I(p)$. If the incumbent does not enter the auction in the first stage and the poacher does, then it is optimal for the incumbent to decline entry in the second stage as well (since he know he will lose the auction). In this case the worker moves to the poacher and receives no transfer. In the case where the incumbent declines to enter in

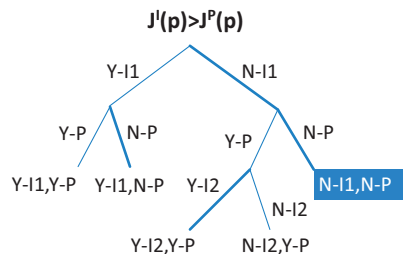


FIGURE A1

Case 1: Incumbent's value with worker greater than poacher's value

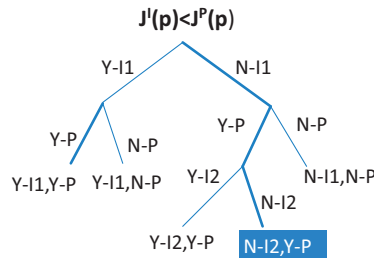


FIGURE A2

Case 2: Incumbent's value with worker less than poacher's value

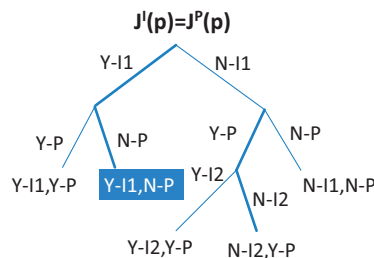


FIGURE A3

Case 3: Incumbent's value with worker equal to poacher's value

the first stage, then the poacher foresees that he will not enter in the second stage either and chooses to pay ε to enter. In the first stage, the incumbent understands that if he enters the auction in the first stage, the poacher will enter and he (the incumbent) will lose the auction. The incumbent therefore chooses not to enter in the first stage. Thus the subgame perfect equilibrium here is (N-I2, Y-P), where the incumbent declines entry in both stages, but the poacher enters and wins over the worker with no transfer.

In the third case ($J^I(p) = J^P(p)$): If the incumbent chooses to enter the auction in the first stage, then it is optimal for the poacher to decline entry, since he knows that he will lose the auction. The worker stays with the incumbent and receives no transfer. Assume the incumbent declines entry in the first stage and the poacher enters. Then it is optimal for the incumbent to decline entry, since if he enters, he pays ε , wins the auction, but gives up his entire share of the surplus by bidding $J^I(p)$. In the case where the incumbent declines to enter in the first stage, it is optimal for the poacher to enter, since he foresees that the incumbent will choose not to enter in the second stage either. In the first stage, it is optimal for the incumbent to enter the auction, since he knows the poacher will not enter. If on the other hand, the incumbent does not enter in the first stage, he knows that the poacher will and the incumbent will not enter in the second stage either. Thus the subgame perfect equilibrium is (Y-I1, N-P), where the incumbent enters in the first stage, but the poacher does not, so there is no transfer to the worker.

Summarizing, a worker who has contacted another firm, always ends up employed in the firm where the surplus is the highest, there is no lump-sum transfer and his wage is given by (1).²¹

21. Mortensen (2003) points out, "unlike in the market for academic economists in the United States, making counteroffers is not the norm in many labour markets. More typically, a worker who informs his employer of a more lucrative outside option is first congratulated and then asked to clear out immediately". Postel-Vinay and Robin (2004) and Moscarini (2008) argue that it may be profitable for firms to commit ex ante not to match outside offers. Burdett and Coles (2003) argue that outside offers are not verifiable and are therefore ignored by the current firm.

Note that the incumbent firm is given two advantages: he has two chances to participate and he wins ties. This, however, is simply a tie-breaking rule which leads the worker to remain at the incumbent in the case of indifference. The model's qualitative features regarding occupational mobility and wages are not affected if ties are broken in a different way.

Note that the worker contacting a firm in the same occupation, corresponds to the third case above ($J^I(p) = J^P(p)$), so the worker stays with the incumbent with no transfer. A worker thus never has an incentive to search for a job in the same occupation.

In our setup, the solution to the Nash bargaining problem results in the linear sharing rule, (2), which provides the necessary condition to determine the worker's wage.

Shimer (2006) has shown that with on-the-job search the linear sharing rule may not always be bilaterally efficient. In particular, he has argued that if the value of the worker is higher in the other occupation, the incumbent employer might have an incentive to pay the worker a higher wage in exchange for not searching on the job. For instance, assume the wage prescribed by the Nash bargaining solution in the other occupation is only slightly higher than the worker's current wage. Then the current firm may find it profitable to increase the worker's wage so that he does not find it profitable to look for a job in the other occupation. In other words, the trade-off faced by the incumbent firm is between a slight reduction in profits, but a discrete jump in the expected duration of its match with the worker, which might make the wage increase optimal. This would in turn imply that the set of feasible payoffs is non-convex, thus violating one of Nash's axioms.

In a framework like the present one, with costless and unobserved job search, such a strategy by the current employer would not work. As Moscarini (2005) notes, if the current employer did offer a higher wage, the worker would have incentive to continue searching on-the-job. If he did contact a firm in the other occupation, the poaching firm can always outbid the incumbent firm in the ensuing auction and offer the worker an (even) higher value. Put differently, changing the current wage does not affect turnover and therefore the duration of the match in the environment studied. Note that this depends on the assumption of costless and unobserved on-the-job search, which rules out this strategy for the firm and preserves the convexity of the set of feasible payoffs. Thus in the present setup, the linear sharing rule (2) is bilaterally efficient.

B. WAGE AND CUTOFFS DERIVATION

The flow value to the firm of a filled vacancy in occupation i satisfies

$$rJ_i(p) = \bar{a}_i(p) - w_i(p) + \frac{1}{2} \zeta_i^2 p^2 (1-p)^2 J_i''(p) - \delta_i J_i(p) - \gamma J_i(p) - \eta_i \lambda J_i(p) I\{V_i(p) < V_k(p)\}, \quad (\text{B.1})$$

where J_i is the asset value of the firm, $\bar{a}_i(p) = p a_i^w + (1-p) a_i^b$ is the expected output as a function of beliefs p , and $I\{\cdot\}$ is an indicator function of whether the worker is searching on the job or not. Therefore, the flow value of the firm is equal to expected output, minus the wage, plus a term that measures the value of learning to the firm, minus the potential capital loss resulting from an exogenous separation, worker death or worker transition to another job. For the firm, unlike the worker, the value of learning is limited only to the duration of the current match. Note that we are assuming that the value of an unfilled vacancy is zero; in the Supplementary Appendix we allow for free firm entry in each occupation, subject to an entry cost, which implies zero vacancy value.

The worker's flow value of being unemployed satisfies

$$rU(p) = b_u + \max_i \lambda (V_i(p) - U(p)) - \gamma U(p), \quad (\text{B.2})$$

i.e. b_u , the value of home production or unemployment benefits, plus the excess value from being employed in occupation i times the job-finding rate, λ , minus the capital loss in case of death. Again note that since workers are learning about their general human capital, the value of being unemployed is a function of the worker's current belief about his type. This is different from Jovanovic (1979) where beliefs are reset upon separation.

Optimal quitting to unemployment for a worker employed in occupation W implies both a value matching ($V_W(p) = U(p)$) or equivalently $J_W(p) = 0$), as well as a smooth pasting condition ($V'_W(p) = U'(p)$) or equivalently $J'_W(p) = 0$). The corresponding conditions for optimal quitting to unemployment in occupation B are

$$V_B(\bar{p}) = U(\bar{p})$$

$$V'_B(\bar{p}) = U'(\bar{p}).$$

As is standard in learning problems, the value functions of the worker and the firm are weakly convex in beliefs p . For the case of an employed worker for example it is true that $E(V_i(p)) \geq V_i(p) = V_i(E(p))$. The equality holds because from the worker's point of view, beliefs are a martingale (if they were expected to go up or down, the worker would not have fully incorporated all available information when updating his prior). The inequality is true because additional information allows the worker to make decisions that improve his situation. For example, if tomorrow's prior decreases, the worker can start searching on the job and achieve higher utility (revealed preference). It therefore follows from Jensen's inequality that V is a convex function of beliefs. The same argument ensures that J_i and U are also convex. The convexity of J_i implies that \underline{p} and \bar{p} are unique.

Furthermore, J_W is globally increasing in p . To understand this, note that $J_W(p) \geq 0 \forall p$, $J_W(\underline{p}) = 0$ and $J'_W(p) = 0$. Since J_W is convex, then it must be increasing everywhere. Intuitively, for \tilde{p} slightly larger than \underline{p} , $J_W(\tilde{p}) > 0$, and thus it must be increasing in that region. Convexity ensures that it is increasing everywhere. Similarly we can show that J_B is decreasing in p .

To see that the threshold \hat{p} is unique, remember that it is defined from $V_W(\hat{p}) - U(\hat{p}) = V_B(\hat{p}) - U(\hat{p})$. However from the Nash bargaining solution we know that $qJ_i(p) = (1-q)[V_i(p) - U(p)]$ which implies that $V_i(p) - U(p) = \frac{q}{1-q}J_i(p)$. Monotonicity and convexity of J_i and therefore of $V_i(p) - U(p)$, implies that $V_W(p) - U(p)$ and $V_B(p) - U(p)$ cross at most once. The assumption that $\lim_{p \rightarrow 0} V_B(p) > \lim_{p \rightarrow 0} V_W(p)$ and that $\lim_{p \rightarrow 1} V_W(p) > \lim_{p \rightarrow 1} V_B(p)$ (a white worker-type is better in a white-collar job, whereas a blue worker-type is more productive in a blue-collar job), ensures that they cross exactly once.

We now derive the solution to the value functions and the cutoffs.

Consider the case where the worker's outside option is the occupation he is currently employed in.

Now equations (3), (B.1), and (B.2) become

$$rV_i(p) = w_i(p) + \frac{1}{2}\xi_i^2 p^2 (1-p)^2 V_i''(p) \quad (3')$$

$$-\delta_i[V_i(p) - U(p)] - \gamma V_i(p)$$

$$rJ_i(p) = \bar{a}_i(p) - w_i(p) + \frac{1}{2}\xi_i^2 p^2 (1-p)^2 J_i''(p) \quad (B.1')$$

$$-\delta_i J_i(p) - \gamma J_i(p)$$

$$rU(p) = b_u + \lambda[V_i(p) - U(p)] - \gamma U(p). \quad (B.2')$$

We subtract the worker's flow value of being unemployed (equation (B.2')) from his flow value of being employed (equation (3')) and multiply through by $(1-q)$

$$\begin{aligned} (1-q)r(V_i(p) - U(p)) &= (1-q)(w_i(p) - b_u) + \frac{1}{2}(1-q)\xi_i^2 p^2 (1-p)^2 V_i''(p) \\ &\quad - \delta_i(1-q)[V_i(p) - U(p)] - \gamma(1-q)V_i(p) \\ &\quad - \lambda(1-q)[V_i(p) - U(p)] + \gamma(1-q)U(p). \end{aligned}$$

We similarly multiply the flow asset value of a filled vacancy (equation (B.1')) by q

$$\begin{aligned} q r J_i(p) &= q \bar{a}_i(p) - q w_i(p) + \frac{1}{2} q \xi_i^2 p^2 (1-p)^2 J_i''(p) \\ &\quad - q \delta_i J_i(p) - q \gamma J_i(p). \end{aligned}$$

We then subtract the above two equations and using the surplus sharing condition (equation (2)) we obtain

$$\begin{aligned} w_i(p) - (1-q)b_u + \frac{1}{2}(1-q)\xi_i^2 p^2 (1-p)^2 V_i''(p) \\ - \lambda(1-q)[V_i(p) - U(p)] - q \bar{a}_i(p) - \frac{1}{2} q \xi_i^2 p^2 (1-p)^2 J_i''(p) = 0, \end{aligned}$$

and therefore

$$\begin{aligned} w_i(p) &= q \bar{a}_i(p) + (1-q)b_u + q \lambda J_i(p) - \frac{1}{2}(1-q)\xi_i^2 p^2 (1-p)^2 V_i''(p) \\ &\quad + \frac{1}{2} q \xi_i^2 p^2 (1-p)^2 J_i''(p). \end{aligned} \quad (B.3)$$

Using the surplus sharing condition once again

$$\begin{aligned} q J_i''(p) &= (1-q)(V_i''(p) - U''(p)) \\ V_i''(p) &= \frac{q}{1-q} J_i''(p) + U''(p). \end{aligned} \quad (B.4)$$

However from the value of the unemployed worker (eq. (B.2')), we have

$$U(p) = \frac{b_u}{r + \gamma + \lambda} + \frac{\lambda}{r + \gamma + \lambda} V_i(p),$$

and therefore

$$U''(p) = \frac{\lambda}{r+\gamma+\lambda} V_i''(p).$$

Substituting out for $U''(p)$ in (B.4) results in

$$V_i''(p) = \frac{q}{1-q} J_i''(p) + \frac{\lambda}{r+\lambda+\gamma} V_i''(p),$$

and therefore

$$V_i''(p) = \frac{q}{1-q} \frac{r+\gamma+\lambda}{r+\gamma} J_i''(p). \quad (\text{B.5})$$

We can now substitute out for $V_i''(p)$ in (B.3) and obtain the wage as a function of J_i only

$$w_i(p) = q\bar{a}_i(p) + (1-q)b_u + q\lambda J_i(p) - \frac{q\lambda}{2(r+\gamma)} \xi_i^2 p^2 (1-p)^2 J_i''(p).$$

Using (B.3) above and the surplus sharing condition (equation (2)), we can derive a more economically intuitive expression for the wage

$$\begin{aligned} w_i(p) &= (1-q)(b_u + \lambda[V_i(p) - U(p)]) + q\bar{a}_i(p) \\ &\quad + \frac{1}{2} \xi_i^2 p^2 (1-p)^2 [q(J_i''(p) + V_i''(p)) - V_i''(p)]. \end{aligned}$$

The case where the worker's occupation of choice if unemployed differs from his current one, *i.e.* $i \neq k$ where i is the worker's current occupation and $k \equiv \arg \max_j [W_j(p) - U(p)]$, is somewhat more involved and we analyse it in the Supplementary Appendix to this article. The wage expression in that case is

$$\begin{aligned} w_i(p) &= (1-q)(b_u + \lambda[V_k(p) - U(p)]) + q\bar{a}_i(p) \\ &\quad + \frac{1}{2} \xi_i^2 p^2 (1-p)^2 [q(J_i''(p) + V_i''(p)) - V_i''(p)] \\ &\quad - \eta_i \lambda [(1-q)(V_k(p) - V_i(p)) + qJ_i(p)], \end{aligned} \quad (\text{B.6})$$

where $k \in \{W, B\}$ and $k \neq i$.

In the first case, the worker's wage weights his outside option and his inside option with weights given by his bargaining power coefficient, q . His outside option consists of his flow utility when unemployed, b_u , plus the option value of search while unemployed. His inside option is his share of the match output plus his share of the match's total value of learning that is *in excess* of his own private value. If $q(J_i''(p) + V_i''(p)) - V_i''(p) < 0$, the worker compensates his employer for the additional benefit he enjoys from learning, by accepting a lower wage. Indeed if we substitute the wage function in the worker's value while employed in occupation $i \in \{W, B\}$ (equation (3')) we get

$$\begin{aligned} rV_i(p) &= (1-q)(b_u + \lambda[V_i(p) - U(p)]) \\ &\quad + q\left(\bar{a}_i(p) + \frac{1}{2} \xi_i^2 p^2 (1-p)^2 (J_i''(p) + V_i''(p))\right) \\ &\quad - \delta_i [V_i(p) - U(p)] - \gamma V_i(p). \end{aligned}$$

This expression reveals that the worker benefits only from his bargained share of the value of learning.

In the second case, where the worker is searching on-the-job, the interpretation is similar, except that the worker's outside option is now different (if unemployed he looks for a job in another occupation). Moreover his wage is reduced by an amount proportional to his search intensity. When the worker leaves his current firm for a firm in another occupation, the separation is no longer bilaterally efficient, as there are lost rents for the incumbent firm. Therefore, when the worker searches on the job, he compensates his firm by an amount equal to the weighted average of the worker's gains, $V_k(p) - V_i(p)$, and the firm's losses, $J_i(p)$, multiplied by his job finding probability. This generates a discontinuity in the wage function.

Substituting for the wage in the firm's value function results in a differential equation with respect to J_i in the case in which the worker's outside option is searching for another job in his current occupation

$$(r+\gamma+\delta_i+q\lambda)J_i(p) = (1-q)(\bar{a}_i(p) - b_u) + \frac{r+\gamma+q\lambda}{2(r+\gamma)} \xi_i^2 p^2 (1-p)^2 J_i''(p), \quad (\text{B.7})$$

and in the case in which it is not

$$\begin{aligned} (r+\gamma+\delta_i+\eta_i\lambda)J_i(p) &= (1-q)(\bar{a}_i(p) - b_u) + \frac{q\lambda}{2(r+\gamma)} \xi_i^2 p^2 (1-p)^2 J_k''(p) \\ &\quad + \frac{1}{2} \xi_i^2 p^2 (1-p)^2 J_i''(p) - q\lambda(1-\eta_i)J_k(p). \end{aligned} \quad (\text{B.8})$$

The general solution to differential equation (B.7) is

$$J_i(p) = \frac{(1-q)(\bar{a}_i(p)-b_u)}{r+\gamma+\delta_i+q\lambda} + K_1^i p^{\frac{1}{2}-\frac{1}{2}\sqrt{\frac{4+h_i}{h_i}}} (1-p)^{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{4+h_i}{h_i}}} \\ + K_2^i p^{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{4+h_i}{h_i}}} (1-p)^{\frac{1}{2}-\frac{1}{2}\sqrt{\frac{4+h_i}{h_i}}},$$

where $h_i = \frac{1}{2} \frac{r+\gamma+q\lambda}{(r+\gamma+\delta_i+q\lambda)(r+\gamma)} \zeta_i^2$ and K_1^i and K_2^i are undetermined coefficients. For the case of $i=W$, when $p \rightarrow 1$, $\lim_{p \rightarrow 1} K_2^W p^{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{4+h_W}{h_W}}} (1-p)^{\frac{1}{2}-\frac{1}{2}\sqrt{\frac{4+h_W}{h_W}}} = K_2^W \cdot 1 \cdot \lim_{p \rightarrow 1} (1-p)^{\frac{1}{2}-\frac{1}{2}\sqrt{\frac{4+h_W}{h_W}}} = +\infty$ which follows from $h_W > 0$, and therefore $\sqrt{\frac{4+h_W}{h_W}} > 1$ and $\frac{1}{2} \left(1 - \sqrt{\frac{4+h_W}{h_W}}\right) < 0$. However since the profits of the firm are bounded from above by the total value of the surplus when the worker is known to be type white, which is finite, it must be the case that $K_2^W = 0$. A similar argument for $i=B$ and $p \rightarrow 0$ leads to $K_1^B = 0$.

Now we can substitute for $J_k(p)$ in (B.8), and using the conditions $J_W(\underline{p})=0$ and $J'_W(\underline{p})=0$ for occupation W ($J_B(\bar{p})=0$ and $J'_B(\bar{p})=0$ for occupation B), we are able to solve the resulting differential equations for the case where the worker is searching on-the-job. The Supplementary Appendix to this article contains the details. The resulting expressions for the asset value of the firm are the following.

For occupation W and $p < \hat{p}$

$$J_W(p) = \frac{r+\gamma+\delta_W+\eta_W\lambda}{\zeta_W^2 l_{WB}} p^{\frac{1}{2}} (1-p)^{\frac{1}{2}} \cdot \int_{\underline{p}}^p \left[\theta_{WB} \tau + \kappa_{WB} K_2^B \tau^{\xi_{WB}} (1-\tau)^{1-\xi_{WB}} + c_{WB} \right] (\tau(1-\tau))^{-\frac{3}{2}} \\ \cdot \left(\left(\frac{1-\tau}{\tau} \frac{p}{1-p} \right)^{l_{WB}} - \left(\frac{\tau}{1-\tau} \frac{1-p}{p} \right)^{l_{WB}} \right) d\tau, \quad (\text{B.9})$$

where

$$l_{WB} = \sqrt{\frac{1}{4} + \frac{2(r+\gamma+\delta_W+\eta_W\lambda)}{\zeta_W^2}} \\ c_{WB} = \frac{1-q}{r+\gamma+\delta_W+\eta_W\lambda} \left(q\lambda(1-\eta_W) \frac{a_B^b - b_u}{r+\gamma+\delta_B+q\lambda} - a_W^b + b_u \right) \\ \theta_{WB} = \frac{1-q}{r+\gamma+\delta_W+\eta_W\lambda} \left(q\lambda(1-\eta_W) \frac{a_B^w - a_B^b}{r+\gamma+\delta_B+q\lambda} - a_W^w + a_W^b \right) \\ \kappa_{WB} = -\frac{q\lambda}{r+\gamma+\delta_W+\eta_W\lambda} \left(\frac{r+\gamma+\delta_B+q\lambda}{r+\gamma+q\lambda} \left(\frac{\zeta_W}{\zeta_B} \right)^2 - 1 + \eta_W \right) \\ \xi_{WB} = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{4+h_B}{h_B}}.$$

Similarly, for the case of occupation B and $p > \hat{p}$

$$J_B(p) = \frac{r+\gamma+\delta_B+\eta_B\lambda}{\zeta_B^2 l_{BW}} p^{\frac{1}{2}} (1-p)^{\frac{1}{2}} \cdot \int_{\bar{p}}^p \left[\theta_{BW} \tau + \kappa_{BW} K_1^W \tau^{\xi_{BW}} (1-\tau)^{1-\xi_{BW}} + c_{BW} \right] (\tau(1-\tau))^{-\frac{3}{2}} \\ \cdot \left(\left(\frac{1-\tau}{\tau} \frac{p}{1-p} \right)^{l_{BW}} - \left(\frac{\tau}{1-\tau} \frac{1-p}{p} \right)^{l_{BW}} \right) d\tau, \quad (\text{B.10})$$

where

$$\begin{aligned}
 l_{BW} &= \sqrt{\frac{1}{4} + \frac{2(r+\gamma+\delta_B+\eta_B\lambda)}{\zeta_B^2}} \\
 c_{BW} &= \frac{1-q}{r+\gamma+\delta_B+\eta_B\lambda} \left(q\lambda(1-\eta_B) \frac{(a_W^b - b_u)}{r+\gamma+\delta_W+q\lambda} - a_B^b + b_u \right) \\
 \theta_{BW} &= \frac{1-q}{r+\gamma+\delta_B+\eta_B\lambda} \left(q\lambda(1-\eta_B) \frac{a_W^w - a_W^b}{r+\gamma+\delta_W+q\lambda} - a_B^w + a_B^b \right) \\
 \kappa_{BW} &= -\frac{q\lambda}{r+\gamma+\delta_B+\eta_B\lambda} \left(\frac{r+\gamma+\delta_W+q\lambda}{r+\gamma+q\lambda} \left(\frac{\zeta_B}{\zeta_W} \right)^2 - 1 + \eta_B \right) \\
 \xi_{BW} &= \frac{1}{2} - \frac{1}{2} \sqrt{\frac{4+h_W}{h_W}}.
 \end{aligned}$$

To complete the solution of the value functions and the worker's decision rules we need to pin down the value of the 5 remaining unknowns: the 3 cutoffs, p , \bar{p} and \hat{p} , as well as the 2 yet undetermined coefficients K_1^W and K_2^B . We need 5 conditions to do so. Optimality of searching behavior while unemployed, $V_W(\bar{p}) = V_B(\bar{p})$, provides one of these conditions. In the Supplementary Appendix we show that the remaining 4 conditions are given by continuity of the total value of the match ($V_i + J_i$) at \hat{p} , as well as continuity of its first derivative, for each occupation. Straightforward derivations imply that they can be rewritten as follows

$$\lim_{p \rightarrow \bar{p}^-} J_W(p) = \lim_{p \rightarrow \bar{p}^+} J_W(p) \quad (\text{B.11})$$

$$\lim_{p \rightarrow \bar{p}^-} J_B(p) = \lim_{p \rightarrow \bar{p}^+} J_B(p) \quad (\text{B.12})$$

$$J_W(\hat{p}) = J_B(\hat{p}) \quad (\text{B.13})$$

$$\lim_{p \rightarrow \bar{p}^-} (V'_W(p) + J'_W(p)) = \lim_{p \rightarrow \bar{p}^+} (V'_W(p) + J'_W(p)) \quad (\text{B.14})$$

$$\lim_{p \rightarrow \bar{p}^-} (V'_B(p) + J'_B(p)) = \lim_{p \rightarrow \bar{p}^+} (V'_B(p) + J'_B(p)). \quad (\text{B.15})$$

The solution of the above non-linear system of 5 equations and 5 unknowns allows us to fully characterize both the wage for every occupation and value of the posterior, as well as the optimal behavior of the worker. As discussed in the Supplementary Appendix, the resulting solution is unique.

C. STATIONARY DISTRIBUTION

Since the process that characterizes the evolution of the beliefs is Markovian and positive recurrent, it has a unique stationary distribution. Let $F_i(p)$ denote the population of workers employed in occupation i whose posterior probability of being white type is less than p . Let $f_i(p)$ denote the corresponding population density of employed workers in occupation i . Similarly, let $Z_i(p)$ be the population of those unemployed and looking for a job in occupation i , whose posterior probability of being a white type is less than p . $z_i(p)$ denotes the corresponding population density of unemployed workers in occupation i .

Following Karlin and Taylor (1981) (chapter 15), the Kolmogorov forward equation for occupation W for every $p \geq \bar{p}$ and $p \neq \hat{p}$ is given by

$$\begin{aligned}
 0 &= \frac{df_W(p)}{dp} = \frac{d^2}{dp^2} \left[\frac{1}{2} \zeta_W^2 p^2 (1-p)^2 f_W(p) \right] - \delta_W f_W(p) - \gamma f_W(p) + \lambda z_W(p) \\
 &\quad - \eta_W \lambda f_W(p) I\{p < \hat{p}\} + \eta_B \lambda f_B(p) I\{\hat{p} \leq p \leq \bar{p}\},
 \end{aligned} \quad (\text{C.16})$$

where $I\{\cdot\}$ is the indicator function.

This equation ensures that flows in and out of employment in occupation W , for every $p \neq \bar{p}$, are equal. The first term captures the net change in p caused by workers moving into p from the right and left of that point, as well as those workers moving away from p . The second and third term measure the outflow from p resulting from exogenous job destruction and worker death shocks respectively. The fourth term captures the inflow of new workers from unemployment at p . Finally, the last two terms reflect the outflow of workers from p to other jobs in occupation B and the worker inflow from

occupation B to newly created matches in occupation W at p . Similarly, the Kolmogorov forward equation for occupation B for every $p \leq \bar{p}$ and $p \neq \underline{p}$ is given by

$$0 = \frac{df_B(p)}{dt} = \frac{d^2}{dp^2} \left[\frac{1}{2} \zeta_B^2 p^2 (1-p)^2 f_B(p) \right] - \delta_B f_B(p) - \gamma f_B(p) + \lambda z_B(p) - \eta_B \lambda f_B(p) I\{\widehat{p} \leq p\} + \eta_W \lambda f_W(p) I\{\underline{p} \leq p < \widehat{p}\}. \quad (C.17)$$

In order to solve for f_W and f_B we first need to solve for the population density of unemployed workers in each occupation $i \in \{W, B\}$, z_i . To do so, we use the fact that in the steady-state, flows in and out of every p in the distributions of unemployed workers must be equal. In particular, the following holds for unemployed workers looking for employment in W , for every $p \geq \widehat{p}$ and $p \neq \bar{p}$

$$\delta_W f_W(p) + \delta_B f_B(p) + \gamma g(p) = \lambda z_W(p) + \gamma z_W(p). \quad (C.18)$$

The first two terms on the left hand side represent the inflow into unemployment due to exogenous separation shocks from occupations W and B , respectively. The third represents the inflow of newly born workers at p . The two terms on the right-hand side account for the exit of workers from p because they either find a job or die. Furthermore, the corresponding condition for unemployed workers in occupation B , is that for every $p < \widehat{p}$ and $p \neq \underline{p}$

$$\delta_W f_W(p) + \delta_B f_B(p) + \gamma g(p) = \lambda z_B(p) + \gamma z_B(p). \quad (C.19)$$

We then use (C.18) and (C.19) to solve out for $z_W(p)$ and $z_B(p)$ respectively and after substituting them into the two forward equations (C.16) and (C.17), we derive the following system of differential equations

$$\begin{aligned} & \frac{d^2}{dp^2} \left[\frac{1}{2} \zeta_W^2 p^2 (1-p)^2 f_W(p) \right] - \gamma \frac{\delta_W + \lambda + \gamma}{\lambda + \gamma} f_W(p) + \frac{\lambda \delta_B}{\lambda + \gamma} f_B(p) \\ & + \frac{\gamma \lambda}{\lambda + \gamma} g(p) - \eta_W \lambda f_W(p) I\{p < \widehat{p}\} + \eta_B \lambda f_B(p) I\{\widehat{p} \leq p \leq \bar{p}\} = 0 \\ & \frac{d^2}{dp^2} \left[\frac{1}{2} \zeta_B^2 p^2 (1-p)^2 f_B(p) \right] - \gamma \frac{\delta_B + \lambda + \gamma}{\lambda + \gamma} f_B(p) + \frac{\lambda \delta_W}{\lambda + \gamma} f_W(p) \\ & + \frac{\gamma \lambda}{\lambda + \gamma} g(p) - \eta_B \lambda f_B(p) I\{\widehat{p} \leq p\} + \eta_W \lambda f_W(p) I\{\underline{p} \leq p < \widehat{p}\} = 0. \end{aligned}$$

Taking cases we are able to solve the above system. We make use of the following conditions to pin down the undetermined coefficients

$$\begin{aligned} & \lambda \int_{\bar{p}}^{\bar{p}} z_W(p) dp + \lambda \int_{\bar{p}}^1 z_W(p) dp + \lambda z_W(\bar{p}) + \eta_B \lambda \int_{\bar{p}}^{\bar{p}} f_B(p) dp \\ & = (\delta_W + \gamma) \int_{\underline{p}}^1 f_W(p) dp + \frac{1}{2} \zeta_W^2 \underline{p}^2 (1-\underline{p})^2 f'_W(\underline{p}) + \eta_W \lambda \int_{\underline{p}}^{\widehat{p}} f_W(p) dp \\ & \int_{\bar{p}}^1 f_W(x) dx \leq 1 < \infty \\ & f_W(\underline{p}) = 0 \\ & \lim_{\varepsilon \rightarrow 0} f_W(\bar{p} + \varepsilon) = f_W(\bar{p}) \\ & \lim_{\varepsilon \rightarrow 0} f_W(\widehat{p} + \varepsilon) = f_W(\widehat{p}) \\ & \lim_{\varepsilon \rightarrow 0} f'_W(\widehat{p} + \varepsilon) = f'_W(\widehat{p}), \end{aligned}$$

and symmetrically 6 more conditions for occupation B . The first condition states that flows in and out of employment in occupation W must equal in the steady state. The first three terms on the left-hand side, capture the inflow of workers from unemployment into employment, at rate λ . At $p = \bar{p}$, the endogenous exit of workers from occupation B , generates an atom of unemployed workers at that point. The fourth term captures the inflow of workers directly from occupation B . The first two terms on the right hand side denote the exit of workers due to endogenous match destruction and death, respectively. The third term captures the endogenous exit of workers at $p = \underline{p}$, while the last term accounts for employed workers who find a job in occupation B .

The second condition states that the mass of workers at any interval is bounded by one, while the third condition is a boundary condition given by optimal worker quitting at p . The remaining three conditions state that the beliefs distribution should be continuous at \bar{p} and \hat{p} and also that that first derivative should be continuous at \hat{p} .

The detailed derivation of the solution is available in the Supplementary Appendix. We verify our solution for the steady state distribution of beliefs by simulating our model and comparing the resulting steady state distribution with the one derived using the equations below.

The resulting steady-state distribution of workers employed in occupation B is the following:

In the case where $p \in [0, \underline{p}]$

$$f_B(p) = C_2^B p^{-1-q_B} (1-p)^{q_B-2} - \frac{d_B}{\sqrt{c_B(4+c_B)}} p^{q_B-2} (1-p)^{-1-q_B} \int_0^p \tau^{\psi_1-q_B} (1-\tau)^{q_B+\psi_2-1} d\tau + \frac{d_B}{\sqrt{c_B(4+c_B)}} p^{-1-q_B} (1-p)^{q_B-2} \int_0^p \tau^{q_B+\psi_1-1} (1-\tau)^{\psi_2-q_B} d\tau, \quad (C.20)$$

for $p \in (p, \hat{p})$

$$f_B(p) = C_3^B p^{q_B-2} (1-p)^{-1-q_B} + C_4^B p^{-1-q_B} (1-p)^{q_B-2} + \frac{1}{\sqrt{c_B(4+c_B)}} p^{q_B-2} (1-p)^{-1-q_B} [C_6^W m_B \int_p^{\hat{p}} \tau^{-\frac{1}{2}-\kappa_B-q_B} (1-\tau)^{-\frac{3}{2}+\kappa_B+q_B} d\tau - C_6^W n_B \int_p^{\hat{p}} \tau^{-\frac{1}{2}+\kappa_B-q_B} (1-\tau)^{-\frac{3}{2}-\kappa_B+q_B} d\tau + d_B \int_p^{\hat{p}} \tau^{\psi_1-q_B} (1-\tau)^{\psi_2+q_B-1} d\tau] - \frac{1}{\sqrt{c_B(4+c_B)}} p^{-1-q_B} (1-p)^{q_B-2} [C_6^W m_B \int_p^{\hat{p}} \tau^{-\frac{3}{2}-\kappa_B+q_B} (1-\tau)^{-\frac{1}{2}+\kappa_B-q_B} d\tau - C_6^W n_B \int_p^{\hat{p}} \tau^{-\frac{3}{2}+\kappa_B+q_B} (1-\tau)^{-\frac{1}{2}-\kappa_B-q_B} d\tau + d_B \int_p^{\hat{p}} \tau^{\psi_1+q_B-1} (1-\tau)^{\psi_2-q_B} d\tau], \quad (C.21)$$

and for $p \in [\hat{p}, \bar{p})$

$$f_B(p) = C_5^B (p(1-p))^{-\frac{3}{2}} \left(\left(\frac{1-p}{p} \right)^{\kappa_W} - \left(\frac{1-\bar{p}}{\bar{p}} \right)^{2\kappa_W} \left(\frac{p}{1-p} \right)^{\kappa_W} \right), \quad (C.22)$$

where $C_2^B, C_3^B, C_4^B, C_5^B$ and C_6^W are undetermined coefficients and $c_B = \frac{\zeta_B^2(\lambda+\gamma)}{2\gamma(\delta_B+\lambda+\gamma)}$, $q_B = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{4+c_B}{c_B}}$, $d_B = -\frac{\lambda}{(\delta_B+\lambda+\gamma)B(\psi_1, \psi_2)}$, $\kappa_W = \sqrt{\frac{1}{4} + \frac{2(\delta_B+\gamma+\eta_B\lambda)}{\zeta_B^2}}$, $\kappa_B = \sqrt{\frac{1}{4} + \frac{2(\delta_W+\gamma+\eta_W\lambda)}{\zeta_W^2}}$, $n_B = -\lambda \frac{\delta_W+\eta_W\lambda+\eta_W\gamma}{\gamma(\delta_B+\lambda+\gamma)}$, $m_B = -n_B \left(\frac{p}{1-p} \right)^{2\kappa_B}$.

Equations (6A) through (8A) of the Supplementary Appendix give us the steady-state distribution of workers employed in occupation W . As shown in the Supplementary Appendix, the 8 undetermined coefficients above are pinned down by 8 conditions, resulting in an 8x8 linear system of equations.

In the Supplementary Appendix we also show that the distribution of posterior beliefs for occupation W (B) features a fat Pareto-type tail if and only if $\gamma > \zeta_W^2 \frac{\lambda+\gamma}{\delta_W+\lambda+\gamma}$ ($\gamma > \zeta_B^2 \frac{\lambda+\gamma}{\delta_B+\lambda+\gamma}$). We derive here conditions under which the wage distribution also features a fat right tail.

The wage function is given by

$$w_B(p) = q\bar{a}_B(p) + (1-q)b_u - \frac{1}{2}\zeta_B^2 p^2 (1-p)^2 (1-q)V_B''(p) + \frac{1}{2}\zeta_B^2 p^2 (1-p)^2 qJ_B''(p) + q\lambda J_B(p),$$

where

$$J_B(p) = \frac{(1-q)(\bar{a}_B(p) - b_u)}{r + \gamma + \delta_B + q\lambda} + K_2^B p^{\frac{1}{2}} \left(1 + \sqrt{\frac{4+h_B}{h_B}} \right) (1-p)^{\frac{1}{2}} \left(1 - \sqrt{\frac{4+h_B}{h_B}} \right).$$

As p approaches zero, we drop all higher terms and we are left with²²

$$w_B(p) \simeq p[q(a_B^w - a_B^b) + q\lambda(1-q)\frac{a_B^w + a_B^b}{r+\gamma+\delta_B+q\lambda}] + q\lambda(1-q)\frac{a_B^b - b_u}{r+\gamma+\delta_B+q\lambda} + qa_B^b + (1-q)b_u.$$

For the case where $q(a_B^w - a_B^b) + q\lambda(1-q)\frac{a_B^w + a_B^b}{r+\gamma+\delta_B+q\lambda} = 1$, we can write the wage function as $w_B(p) = p + c_0$ where $c_0 = q\lambda(1-q)\frac{a_B^b - b_u}{r+\gamma+\delta_B+q\lambda} + qa_B^b + (1-q)b_u$. In this case the density of wages is given by $f(w - c_0)$, also features a fat *left* tail of the Pareto-type. Alternatively in the case where $q(a_B^w - a_B^b) + q\lambda(1-q)\frac{a_B^w + a_B^b}{r+\gamma+\delta_B+q\lambda} = -1$ the wage distribution in occupation B , features a fat *right* tail of the Pareto-type. Notice that this has an intuitive economic explanation. In both cases there are few workers whose posterior is in the vicinity of zero (as long as $\gamma > \zeta_B^2 \frac{\lambda+\gamma}{\delta_B+\lambda+\gamma}$). The first case is the high-ability, low-ability workers case ($a_B^w - a_B^b = \frac{r+\gamma+\delta_B+q\lambda}{q(r+\gamma+\delta_B+q\lambda+\lambda(1-q))} > 0$ which implies $a_B^w > a_B^b$). Now these workers receive the lowest wages in occupation B , since they are more likely to be the low-ability ones. Hence the within occupation wage distribution features a fat left tail!

In the alternative case ($a_B^w - a_B^b = -\frac{r+\gamma+\delta_B+q\lambda}{q(r+\gamma+\delta_B+q\lambda+\lambda(1-q))} < 0$ which implies $a_B^w < a_B^b$), these workers receive the highest wages in the occupation as their expected productivity is the highest. Now the within occupation wage distribution features a fat right tail, consistent with empirical observations.

Similarly the wage distribution in occupation W features a fat right tail if $a_W^w - a_W^b = \frac{r+\gamma+\delta_W+q\lambda}{q(r+\gamma+\delta_W+q\lambda+\lambda(1-q))}$.

Acknowledgments. I thank the editor and three anonymous referees for very constructive comments. I also thank Giuseppe Moscarini for his invaluable advice and guidance throughout this project, as well as Joe Altonji, Björn Brügemann and Bill Brainard for their suggestions and encouragement. Moreover I am grateful to Manolis Galenianos and Richard Rogerson for extensive comments and suggestions, as well as to numerous colleagues and seminar participants and especially Costas Arkolakis, Alessandro Bonatti, Eduardo Engel, Nicholas Kalouptsidis, Omar Licandro, Iouri Manovskii, Rafael Lopes de Melo, Tony Smith and Neil Wallace. The kind hospitality of the Department of Economics at Princeton while working on a revision is gratefully acknowledged.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- ABOWD, J. M., KRAMARZ, F. and MARGOLIS, D. N. (1999), "High Wage Workers and High Wage Firms", *Econometrica*, **67**, 251–333.
- ACEMOGLU, D. and SHIMER, R. (1999), "Efficient Unemployment Insurance", *Journal of Political Economy*, **107**, 893–928.
- ALTONJI, J. G. (2005), "Employer Learning, Statistical Discrimination and Occupational Attainment", *American Economic Review*, **95**, 112–117.
- ALTONJI, J. G. and PIERRET, C. (2001), "Employer Learning and Statistical Discrimination", *Quarterly Journal of Economics*, **116**, 313–350.
- ANTONOVICS K. and GOLAN, L. (2012), "Experimentation and Job Choice", *Journal of Labor Economics*, **30**, 333–366.
- BAGGER, J., FONTAINE, F., POSTEL-VINAY, F. and ROBIN, J.-M. (2013), "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics", *American Economic Review*, forthcoming.
- BURDETT K. and COLES, M. (2003), "Equilibrium Wage-Tenure Contracts", *Econometrica*, **71**, 1377–1404.
- CAHUC, P., POSTEL-VINAY, F. and ROBIN, J.-M. (2006), "Wage Bargaining with On-the-Job Search: Theory and Evidence", *Econometrica*, **74**, 323–364.
- EECKHOUT, J. (2006), "Employer Learning and General Human Capital" (Mimeo, University of Pennsylvania, Department of Economics).
- EECKHOUT, J. and WENG, X. (2010), "Assortative Learning" (Mimeo, University of Pennsylvania, Department of Economics).
- FARBER, H. S. and GIBBONS, R. (1996), "Learning and Wage Dynamics", *Quarterly Journal of Economics*, **111**, 1007–1047.
- FELLI L. and HARRIS, C. (1996), "Learning, Wage Dynamics, and Firm-Specific Human Capital", *Journal of Political Economy*, **104**, 838–868.

22. Formally, since $\frac{1}{2} \left(1 + \sqrt{\frac{4+h_B}{h_B}}\right) > 1$, there exists a $p_* > 0$ (which is a function of all the parameters in the wage equation), such that for $p \leq p_*$, the equation below holds.

- FLINN, C. (2006), "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contract Rates", *Econometrica*, **74**, 1013–1062.
- FRENCH, E. and TABER, C. (2011), "Identification of Models of the Labor Market", in Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, Vol. 4a (Amsterdam: Elsevier Science B. V.) 537–617.
- GATHMANN, C. and SCHÖNBERG, U. (2010), "How General Is Human Capital? A Task-Based Approach", *Journal of Labor Economics*, **28**, 1–49.
- GIBBONS, R., KATZ, L. F., LEMIEUX, T. and PARENT, D. (2005), "Comparative Advantage, Learning and Sectoral Wage Determination", *Journal of Labor Economics*, **23**, 681–723.
- GORRY, A. (2011), "Experience and Worker Flows" (Mimeo, University of California, Santa Cruz, Department of Economics).
- GROES, F., KIRCHER, P. and MANOVSKII, I. (2010), "The U-Shapes of Occupational Mobility" (Mimeo, University of Pennsylvania, Department of Economics).
- DEN HAAN, W. J., RAMEY, G. and WATSON, J. (2000), "Job Destruction and Propagation of Shocks", *American Economic Review*, **90**, 482–498.
- HARRIS, M. and HOLMSTROM, B. (1982), "A Theory of Wage Dynamics", *The Review of Economic Studies*, **49**, 315–333.
- HECKMAN, J. J. and HONORÉ, B. E. (1990), "The Empirical Content of the Roy Model", *Econometrica*, **58**, 1121–1149.
- HILL, D. H. (1994), "The Relative Empirical Validity of Dependent and Independent Data Collection in a Panel Survey", *Journal of Official Statistics*, **10**, 359–380.
- JOVANOVIĆ, B. (1979), "Job Matching and the Theory of Turnover", *Journal of Political Economy*, **87**, 972–990.
- JOVANOVIĆ, B. and MOFFITT, R. (1990), "An Estimate of a Sectoral Model of Labor Mobility", *Journal of Political Economy*, **98**, 827–852.
- KAMBOUROV, G. and MANOVSKII, I. (2008), "Rising Occupational and Industry Mobility in the United States: 1968–1997", *International Economic Review*, **49**, 41–79.
- KAMBOUROV, G. and MANOVSKII, I. (2009a), "Occupational Specificity of Human Capital", *International Economic Review*, **50**, 63–115.
- KAMBOUROV, G. and MANOVSKII, I. (2009b), "Occupational Mobility and Wage Inequality", *Review of Economic Studies*, **76**, 731–759.
- KARLIN, S. and TAYLOR, H. M. (1981), *A Second Course in Stochastic Processes* (New York: Academic Press).
- LANGE, F. (2006), "The Speed of Employer Learning", *Journal of Labor Economics*, **25**, 1–35.
- LEE, D. and WOLPIN, K. I. (2006), "Intersectoral Labor Mobility and the Growth of the Service Sector", *Econometrica*, **74**, 1–46.
- LIPTSER, R. and SHYRYAEV, A. (1977), *Statistics of Random Processes*, Vol. 2 (Berlin: Springer-Verlag).
- LOPES DE MELO, R. (2007), "The Implications of Search Models for Wage Dynamics: an Empirical Assessment" (Mimeo, Yale University, Department of Economics).
- MACDONALD, G. M. (1982), "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information", *The American Economic Review*, **72**, 1038–1055.
- MARIMON, R. and ZILIBOTTI, F. (1999), "Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits", *The Economic Journal*, **109**, 266–291.
- MCCALL, B. (1990), "Occupational Matching: A Test of Sorts", *Journal of Political Economy*, **98**, 45–69.
- MILLER, R. A. (1984), "Job Matching and Occupational Choice", *Journal of Political Economy*, **92**, 1086–1120.
- MORTENSEN (2003), *Wage Dispersion: Why are Similar Workers Paid Differently?* (Cambridge, MA: MIT Press).
- MOSCARINI, G. (2001), "Excess Worker Reallocation", *Review of Economic Studies*, **68**, 593–612.
- MOSCARINI, G. (2005), "Job Matching and the Wage Distribution", *Econometrica*, **73**, 481–516.
- MOSCARINI, G. (2008), "Job-to-Job Quits and Corporate Culture" (Mimeo, Yale University, Department of Economics).
- MURPHY, K. and TOPEL, R. (1987), "The Evolution of Unemployment in the United States: 1968–1985", in Fischer, S. (ed.), *NBER Macroeconomic Annual 1987* (Cambridge, MA: MIT Press).
- NEAL, D. (1999), "The Complexity of Job Mobility among Young Men", *Journal of Labor Economics*, **17**, 237–261.
- VAN OURS, J. C. and VODOPIVEC, M. (2008), "Does Reducing Unemployment Insurance Generosity Reduce Job Match Quality?", *Journal of Public Economics*, **92**, 684–695.
- PISSARIDES, C. (1994), "Search Unemployment with On-the-job Search", *Review of Economics Studies*, **61**, 457–475.
- POSTEL-VINAY, F. and ROBIN, J.-M. (2002), "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity", *Econometrica*, **70**, 2295–2350.
- POSTEL-VINAY, F. and ROBIN, J.-M. (2004), "To Match or Not to Match? Optimal Wage Policy with Endogenous Worker Search Intensity", *Review of Economic Dynamics*, **7**, 297–330.
- ROY, A. D. (1951), "Some Thought on the Distribution of Earnings", *Oxford Economics Papers*, **3**, 135–146.
- SANDERS, C. (2010), "Skill Uncertainty, Skill Accumulation, and Occupational Choice" (Mimeo, University of Wisconsin, Madison, Department of Economics).
- SCHÖNBERG, U. (2007), "Testing for Asymmetric Employer Learning", *Journal of Labor Economics*, **25**, 651–691.
- SHIMER, R. (2006), "On-the-Job Search and Strategic Bargaining", *European Economic Review*, **50**, 811–830.
- SHIMER, R. (2012), "Reassessing the Ins and Outs of Unemployment", *Review of Economic Dynamics*, **15**, 127–148.