Learning Your Comparative Advantages

Theodore Papageorgiou*

Penn State University, Department of Economics†

September 2012

Abstract

In this paper, we argue that worker learning about their unobserved abilities can account for several key facts that characterize occupational mobility. While employed, workers learn about these skills and self-select into the occupation that best matches their abilities. Our setup is an equilibrium Roy (1951) model with learning that can explain the excess of gross occupational mobility over the net, as well as the decline in occupational mobility with age. In addition, it is consistent with the experience profile of wages and the cross-sectional within-occupation wage inequality among observationally similar workers. The calibration of our framework favors a setup in which each worker is the most productive in some occupation, rather than the hierarchical model of ability in which some workers perform better in all occupations. Finally, we investigate the interaction of information and search frictions and illustrate how an increase in unemployment benefits leads to a reduction in output per worker due to reduced learning.

Keywords: Occupational mobility, Learning, Unemployment

JEL Classification: E24, E25, J24, J62

*I thank the editor and three anonymous referees for their comments and suggestions. 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, Iourii Manovskii, Rafael Lopes de Melo, Tony Smith and Neil Wallace.

†418 Kern Building, University Park, PA 16802, USA. E-mail: tpapageorgiou@psu.edu. Web: http://sites.google.com/site/theodorepapageorgiou/
1 Introduction

The assignment of workers to the tasks or occupations in which they perform best is a fundamental issue in economics with obvious productivity implications. A key feature of occupational choice is that it is not permanent: every year a significant fraction of workers switches occupations.\(^1\) Moreover, gross worker flows across occupations, vastly exceed net flows,\(^2\) suggesting that aggregate shocks to the supply or demand of a given occupation can, potentially account for only a small fraction of occupational flows. Finally, younger workers are significantly more likely to switch\(^3\) and occupational mobility does not appear to be random, i.e. from a given occupation, transitions to some occupations are more likely than to others.\(^4\)

This paper demonstrates, both qualitatively and quantitatively, that a simple model that assigns a key role to learning about occupational comparative advantage is consistent with these key facts.

In our model, each worker is endowed with a type and different worker-types exhibit different productivity levels at different occupations; a special case of our setup is the hierarchical ability model where one worker-type performs better in all occupations. Each worker, however, does not observe his type. He only gradually learns it by observing his output at a given occupation, which is a noisy signal of his type-specific productivity. Since the information accumulated in any occupation is relevant to the choice of the next one, overall labor market experience, rather than tenure as is conventional, matters for expected productivity and wages.\(^5\) Unlike models where the worker learns about the latent quality of his current match only, in the current setup beliefs do not reset once a worker separates from his match; this persistence of beliefs complicates both the solution to the worker’s problem, as well as the derivation of the cross-sectional distribution of beliefs.

Our setup can be viewed as an equilibrium Roy model with learning and unlike the standard Roy model, which forms the benchmark model for occupational choice, it is consistent with the key facts outlined above. First, the model is consistent with the observed offsetting flows across occupations: workers exit a given occupation as they realize they

\(^1\) 13% of workers switch one-digit occupations every year according to Kambourov and Manovskii (2008).
\(^2\) Kambourov and Manovskii (2008), Jovanovic and Moffitt (1990) and Murphy and Topel (1987)
\(^3\) Neal (1999), Kambourov and Manovskii (2008)
\(^4\) See Table 1 in Section 4 and related discussion.
\(^5\) For instance, a worker employed as a sales associate might realize that he possesses good communication skills and switch to a career in advertising.
are more productive elsewhere, while simultaneously other workers enter that occupation, as they believe it to be a good fit for them. Moreover, young workers, who have little labor market experience, are the least informed about their type, are less likely to choose the occupation in which they perform best and are thus more likely to switch, consistent with the observed negative relationship between age and the occupational switching probability. In addition, since workers make occupational choices based on their past experience, occupational choice is not random, i.e. there exist regularities among the occupational transition patterns, as seen for instance in Table 1 of Section 4. Finally, our setup is consistent with a few other important facts regarding occupational choices and labor market dynamics. In particular, it is consistent with return mobility, whereby a fraction of switchers return to their original occupation.\(^6\) In our framework workers can return to their original occupation, either because they have revised upwards their beliefs regarding the original occupation or because they have revised their beliefs downwards regarding their new occupation, which earlier appeared to be a better match than the original one. Moreover our framework illustrates a mechanism that generates occupational mobility with wage cuts:\(^7\) in our setup, different occupations may imply a different speed at which workers learn about their type; workers are therefore willing to work for a lower wage as long as they are compensated by a corresponding increase in the speed of learning. In addition, our model is consistent with two other well-known labor market facts: the increase of wages over the worker’s lifetime (returns to experience) and the cross-sectional wage inequality among observationally similar workers.

We assess the model’s ability to account for the key stylized facts of occupational choice in two complimentary ways. First, in a two-occupation two-type version of the model we are able to fully characterize the equilibrium and list its properties. We then offer a more quantitative assessment of the model’s ability to account for these facts by generalizing our setup to allow for three occupations, a continuum of types and human capital accumulation and using the 1996 panel of the Survey of Income and Program Participation (SIPP). Besides assessing the ability of our relatively simple model to replicate key features of the data, the goal of the quantitative exercise is to uncover the importance of the learning mechanism, as well as obtain estimates of the relative productivity distribution. We are also interested in using the parameters from the quantitative exercise to

\(^6\)30% of workers who switch return to their one-digit occupation within a 4-year period according to Kambourov and Manovskii (2008).
\(^7\)For evidence of job-to-job mobility involving wage cuts see Postel-Vinay and Robin (2002) and Lopes de Melo (2007).
carry out policy exercises.\footnote{Our exercise is partly motivated by the results of Farber and Gibbons (1996) and Altonji and Pierret (2001) who find that the market learns over time about workers’ unobserved characteristics. In particular, Farber and Gibbons (1996) show that the correlation between a worker’s unobserved characteristics (proxied by test scores) and his wage should increase over time, as learning accumulates. Altonji and Pierret (2001) find that the correlation between a worker’s observable characteristics (such as education) and wages should fall over time, as the market learns about the unobserved productivity directly, by observing output. Guvenen’s (2007) results are also consistent with our proposition; using consumption data, he finds that a framework in which workers face individual-specific income profiles and learn about their shape fits the data better than one in which workers are identical but subject to large and very persistent shocks.}

Our setup is calibrated to match the wages of occupational choices of experienced workers who have learned their type and those entering the labor market. Even though we allow for the hierarchical model of ability in which some workers perform better in all occupations, our results favor a setup in which each worker-type is the most productive in some occupation. In particular, the hierarchical ability model predicts that there is very little overlap in the support of the wage distributions of different occupations, whereas in the data the within-occupation wage distributions are fairly similar. This result is reminiscent of Willis and Rosen (1979), who also find evidence against the hierarchical ability model in a very different context, namely the decision to attend college.

Using the parameters from the calibrated model, we explore the aggregate implications of this type of learning. Our setup illustrates a novel channel through which unemployment is costly. In particular, the rate at which workers learn about their abilities affects the allocation of labor across 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 rather than employed and learning about their abilities and are on average more uncertain about their skills. We also calculate the welfare-maximizing level of unemployment benefits.

The paper is organized as follows: the next section discusses the paper’s relation to the literature. Section 3 presents the two-occupation two-type version of the model and characterizes the equilibrium; Section 4 discusses model implications; Section 5 contains the quantitative exercise which is based on the more general model. We offer some concluding remarks in Section 6. The appendix, as well as the online appendix, available on the author’s website, contain detailed derivations and proofs.
2 Relation to the Literature

This work is related to a number of papers which can be roughly categorized into those focusing on occupations and occupational mobility, those focusing on match specific learning and those focusing on workers learning about their type.

The first group of papers demonstrates the recent growing interest in theories of occupation specific human capital. For instance, Kambourov and Manovskii (2009a) find that returns to occupational tenure are significant, in contrast to employer tenure. In a related paper, Kambourov and Manovskii (2009b) argue that aggregate shocks and occupation-specific human capital are key in explaining the increase in occupational mobility from the 1970s to the 1990s. In their model workers choose to leave occupations that are hit by negative shocks. In the data however, there are large offsetting flows across all occupations: for every worker entering an occupation, there is usually another worker leaving it (even in expanding occupations).\footnote{For instance, using data from the SIPP one observes that 7.44\% of its workforce left the managerial occupational group (003-037) for another occupation every year on average, between 1996 and 1999, even though that occupation expanded its employment share from 13.28\% to 13.81\% during the same period.} In contrast, in our setup, mobility is generated through learning which can naturally explain the large excess of gross flows over the net.

Regarding the second group of papers, a large literature examines the interaction of learning and labor markets where the worker learns about the latent quality of his firm/occupational match (Jovanovic (1979), Miller (1984), McCall (1990), Moscarini (2005)). The key difference in the predictions of this class of models compared to our setup stems from the fact that current match signals are not informative about future matches. As a result, in those papers, workers who separate from their match search for new jobs randomly. Moreover, workers do not return to occupations they have abandoned: even if his match with the new occupation proves to be unsuccessful, the worker prefers to try out a new occupation if one is available. Finally, wage growth depends on occupation tenure rather than overall labor market experience. Although there is a large literature debating the importance of tenure in wage formation, there is ample evidence that labor market experience is a fundamental determinant of wages, consistent with our setup.\footnote{Abraham and Farber (1987), Altonji and Shakotko (1987) and Altonji and Williams (2005) find small effects of firm tenure on the wage, whereas Topel (1991) reports larger estimates. More recently, Kambourov and Manovskii (2009a) present results that underline the importance of occupation rather than firm or industry tenure on wages, while Pavan (2011) finds positive returns for both career and firm tenure. Gathmann and Schönberg (2010) provide evidence that tenure in the previous occupation also affects the current wage, while emphasizing the importance of task tenure in wage formation.}

Within this class of models, Miller (1984) extends the idea of learning about match
quality to occupations.\textsuperscript{11} In addition to the above-mentioned differences in predictions, Miller’s (1984) setup predicts that all workers start off in the same occupation and go through the same succession of occupations, while they switch occupations only after exhausting all available jobs in that occupation. We explore the empirical validity of this prediction using the SIPP and observe that flows across occupations are bidirectional: for most occupational pairs, the flow of workers moving in one direction is of the same order of magnitude as those moving in the opposite direction, contrary to the predictions of Miller.\textsuperscript{12} Furthermore, among workers who switch employers, 76% move on to another occupation, rather than remain in the same one, as Miller’s framework would predict.

Also within this class of models, Moscarini (2005) is the most closely related paper from a methodological perspective. In his work the worker learns about the underlying quality of his match with a firm, which can be either high or low. This binary restriction implies that beliefs follow a Bernoulli distribution, which is also the case in the main model of the present paper, allowing for the derivation of closed form expressions for the value functions, the distribution of beliefs and the wage distributions.

In addition to the differences in predictions that were stressed in the previous paragraph between the current setup and this class of models, in our setup the persistence of beliefs presents additional technical challenges: in Moscarini’s setup, as well as all other models within this class, once a worker separates from his match his beliefs reset. In contrast, here, the worker learns about his own type so that output realizations are informative not only about his current match, but also about other matches and unemployment. This implies for instance, that the value of an unemployed worker now depends on his beliefs. The persistence of the state variable complicates the problem in the following ways: first deriving the equilibrium belief distribution becomes considerably more involved. Because workers with different beliefs may decide to enter an occupation, inflows can occur at a large part of the wage distribution;\textsuperscript{13} in contrast, in Moscarini (2005), all workers enter an occupation with the same prior belief. Second, the solution to the worker’s problem (when to search on-the-job, when to quit to unemployment) now also depends on the continuity of the total value of the match, as well as continuity of its first derivative at the point where the worker begins on-the-job search, in addition to the usual value matching and smooth pasting conditions that characterize his decision to quit to

\textsuperscript{11}He assumes that jobs with similar expected productivity and rate of information acquisition belong to the same occupation.

\textsuperscript{12}See the occupational transition matrix in Section 4 (Table 1).

\textsuperscript{13}The resulting differential equation is more complicated. Moreover, there are more undetermined coefficients and therefore a number of additional conditions is necessary to pin them down (see results in Appendix C, as well as Section 3 of the online appendix).
unemployment.\textsuperscript{14} Furthermore, in our setup, the condition governing whether the belief distribution follows a fat Pareto tail, depends on the death rate (as beliefs are reset only at that point), rather than on the separation rate as in Moscarini (2005). Finally, in the model presented in Section 5.1, we move beyond the binary restriction and allow workers to draw their productivities from a multivariate normal distribution; this more general model is employed in the empirical exercise.

We next turn to the third class of models in which our setup belongs. Harris and Holmstrom (1982), one of the first models on worker learning about their abilities, consider a setup where firms provide insurance to risk-averse workers who don’t know their ability. Their setup focuses mostly on the implications regarding wage growth and downward wage rigidity, whereas our setup also focuses on the implications of learning regarding occupational choice. MacDonald’s (1982) theoretical work shows that learning about one’s ability can reproduce the increase in wages over a worker’s lifetime, as well as a positively skewed wage distribution. His paper’s predictions differ from ours in several respects. First his setup predicts that all workers receive the same wage within an occupation, in contrast to our setup which matches the shape of the within-occupation cross-sectional wage distribution, including its well-known fat right tail; indeed empirical evidence suggests that substantial wage dispersion remains within occupational groups (Hornstein et al. (2007)). Second, in MacDonald’s setup all workers begin their career with the same prior belief and thus choose the same occupation. The model in the present paper, instead allows workers to start their working careers with different priors and therefore make different initial occupational choices, consistent with empirical evidence. Indeed, we estimate a multinomial logit model of occupational choice as in Schmidt and Strauss (1975), for workers with up to two years of potential labor market experience and find substantial differences among the occupational choice of young workers.\textsuperscript{15} Third, our setup is consistent with the decline in the occupational mobility as workers age. Finally, we empirically assess our setup and we incorporate several realistic features: different occupations can have different speeds of learning, which workers take into account in their decisions; we include search frictions; we allow for general human capital accumulation.

Felli and Harris (1996) focuses on the returns to firm-specific human capital and does not have predictions regarding occupational mobility, but shares conceptual similarities with the present paper. Their paper, like ours, considers learning in a continuous time labor market setting. In the present paper’s framework however, (informational) human

\textsuperscript{14}Here as well, the resulting differential equation is more complicated with more unknowns (see results in Appendix B, as well as Section 1 of the online appendix).

\textsuperscript{15}Results available upon request.
capital is more general and transferable across firms: even within an occupation, competition among firms in that occupation ensures that the worker receives at least part of the returns to his human capital. In the case of firm-specific human capital, firm competition does not ensure that the worker always receives part of the returns and this is the problem Felli and Harris tackle. In their setup, in equilibrium, a worker also invests in learning about his productivity in the task valued by another firm as well and his wage largely reflects his expected output in the other firm. On the contrary, in the setup in this paper, the wage largely reflects his expected output in his current occupation.

In recent work, Eeckhout and Weng (2010), consider a setup similar to ours and show that there is always positive assortative matching in equilibrium under strict supermodularity. The main difference with the framework presented here is that ours includes search frictions.\textsuperscript{16}

On the empirical side, Gibbons et al. (2005) develop a setup with worker learning about abilities that are valued differentially across sectors. They estimate their model using past occupational decisions as instruments and find evidence of sorting across sectors: high-wage sectors employ high-skill workers and offer high returns to skill. In a related paper, Groes et al. (2010) document that the workers with the highest and lowest wages within an occupation are the ones most likely to move (low wage workers are more likely to switch to occupations with lower average wages, while high wage workers are more likely to switch to occupations with higher average wages) and then construct a model of worker learning about their unobserved abilities consistent with their findings.

Both of these setups consider a hierarchical setup, with one dimension of skills, so higher ability workers produce more in all occupations; in contrast, our model also allows for an ability distribution where each worker type is the most productive in some occupation. Moreover, occupational choice in the above models largely depends on expected output, whereas in our framework workers maximize their value which, in addition to expected output, depends on the degree of search frictions, which are absent in the above frameworks, and the speed of learning which we allow to differ across occupations. In Section 5.4 we discuss the different predictions of the two setups and how the results from

\textsuperscript{16}They show that in the frictionless environment there is an additional condition for optimal worker switching that restricts the second derivative of the value function.
3 Model and Equilibrium

We begin by analyzing the case where there are two occupations and a worker’s type can take on only one of two values. We focus on this setup because we are able to obtain an analytic characterization and thereby most clearly demonstrate the key economic mechanisms at work. In Section 5 we relax these two assumptions.

The basic environment is the following: there are two occupations, blue collar and white collar; workers are assumed to be of two types, each type being most productive in a different occupation. We identify a worker’s type by the occupation in which they have an advantage, so a “white” type is more productive in a white collar job than a blue collar job and “blue” type is more productive in a blue collar job than a white collar one. In frictionless world, white workers enter the white collar occupation and blue workers enter the blue collar occupation. We, however, introduce two key frictions: first, individuals do not know their type, but learn it over time based on their labor market experience. Second, there are search frictions, so that individuals cannot costlessly obtain employment or switch occupations.

We next describe our setup in detail.

3.1 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 $\gamma$, while new workers are born at the same rate, ensuring 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

---

17 Altonji (2005) argues that if high-skilled jobs exhibit a higher speed of learning, then the market will be slow to learn the productivity of a worker who starts out in a low-skilled job. Similarly, in Antonovics and Golan’s (2012) setup workers can choose to learn about a specific skill and also how much to learn. As in our framework, workers face a trade-off between increased learning about their ability and higher current wages. Both papers assume a hierarchical ability level, a special case of our framework, which, as in Roy (1951), also allows each worker-type to be the most productive in some occupation.

18 Recent work by Sanders (2010) argues that skill uncertainty plays a significant role in explaining worker mobility across different tasks. Gorry (2011) argues that learning from past experience can account for the observed declining in the job finding rate as workers grow older.
unemployed and employed, can search in only one occupation and have to choose which occupation to search in.\textsuperscript{19} Search is costless. An unemployed worker meets firms at an exogenous rate $\lambda$ (we endogenize $\lambda$ in the online 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 $\eta_i > 0$.\textsuperscript{20} Worker-firm matches dissolve exogenously at rate $\delta_i$ (e.g. plant closing) after which the worker becomes unemployed. As shown in the next section, there is also 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, $W$ and $B$.\textsuperscript{21} The flow match output produced from a match $\kappa$, between a worker of type $\tau$, $\tau \in \{w, b\}$, and a firm in occupation $i$, $i \in \{W, B\}$, is determined by:

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

where $Z_t^\kappa$ is a match-specific Wiener process, $a_i^\tau$ is the occupation and type specific mean, whereas $\sigma_i^\tau$ is the occupation-specific output noise.\textsuperscript{22} 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 $B$ ($a_w^W > a_b^W$). Moreover, we assume that type blue ($b$) workers are more productive in occupation $B$, than in occupation $W$ ($a_b^B > a_w^B$).\textsuperscript{23}

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

\textsuperscript{19}For instance, if a worker has one unit of time to search, searching in only one occupation is optimal if the search technology exhibits constant or increasing returns.

\textsuperscript{20}Even if $\eta_i > 1$, i.e. searching on-the-job is more effective than when unemployed, a worker may still find it optimal to quit to unemployment if his expected output is sufficiently below $b_u$. The results of the empirical exercise (Section 5.4) however, are consistent with $\eta_i < 1$ for all occupations.

\textsuperscript{21}It is straightforward to extend the model to more than two occupations, but with still two worker-types.

\textsuperscript{22}One may imagine that the true underlying function that maps worker-type to output is more complicated. If that is the case, then learning becomes more difficult and therefore $\sigma_i$ here captures both output volatility, as well as the (unmodeled) complexity of the production function.

\textsuperscript{23}The case where both types are most productive in the same occupation is not interesting, since then all workers choose that occupation regardless of their beliefs.
are determined by generalized Nash bargaining, with \( q \in (0, 1) \) denoting the worker’s bargaining power.\(^{24}\)

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, \( dY_t^e \), the market and the worker update their belief regarding the worker’s type, \( \tau_i \), 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, as well as his initial belief. As in Liptser and Shyryaev (1977), using Bayes’ rule and Ito’s lemma, the belief process is reduced to:

\[
dp_t = p_t(1 - p_t) \frac{dY_t^e - \left(p_t a_i^w + (1 - p_t) a_i^b\right)}{\sigma_i} dt
\]

where \( \zeta_i \equiv \frac{a_i^w - a_i^b}{\sigma_i} \). The evolution of beliefs depends on the realized match output outcome, \( dY_t^e \), relative to the expected one, \( p_t a_i^w + (1 - p_t) a_i^b \). Depending on whether \( a_i^w \) is greater than or less than \( a_i^b \), (and thus whether \( \zeta_i \) is positive or negative), a surprisingly high output realization updates \( p \) to either a higher or a lower value, meaning that the worker is either more or less likely to be 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, \( \zeta_i \)) and the level of output noise, \( \sigma_i \), separately. A low variance of beliefs signifies that the worker is confident about his type and thus moderately revises his belief as new information arrives. If the performance of the two types in occupation \( i \) is significantly different (\( |a_i^w - a_i^b| \) is large), it is easier to distinguish between the two. Similarly, a low level of output noise (\( \sigma_i \)) implies that more information is revealed about the worker’s true productivity, \( a_i^\tau \), implying a higher speed of learning. Finally, note that \( \sigma_i \) also appears in the last term on the right hand side and scales the size of the news: whether an output realization is “surprising”, depends on the level of output noise in that occupation; a more “surprising” realization contains more information and therefore leads to a larger revision in beliefs. This last term on the right hand side is a standard Wiener process with respect to the unconditional probability measure over types and signal realizations used by the agents of this economy.

Note that as workers update their beliefs, they revise not only their expected per-

\(^{24}\)We are abstracting from any sources of asymmetric information, and view this as a natural benchmark. For an example of a three-period labor market model with firm asymmetric information the interested reader should refer to Eeckhout (2006). See also results in Schönberg (2007).
formance in their current occupation, but also their expected productivity in the other occupation (even if they have never been employed in it).

3.2 Equilibrium

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

3.2.1 Wages

Since \( p \) 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 change about the worker’s type, so does the worker’s wage, since 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 employing 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, as mentioned above, by generalized Nash bargaining, which dictates that a worker’s wage in occupation \( i \in \{W, B\} \), \( w_i(p) \), is set as a function of beliefs \( p \):

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

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 firm place bids in order to attract the worker. Moreover, a worker never switches to another firm in the same occupation, as he is equally valuable to both firms. Thus a worker chooses to search on-the-job only for firms in the other occupation.\(^{25}\)

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

\[
qJ_i(p) = (1 - q) [V_i(p) - U(p)]
\]

\(^{25}\)In principle, workers can also search on-the-job for firms in the same occupation since search is costless, but they gain nothing by meeting another firm in this case (see also Appendix A).
which provides the necessary condition to determine the worker’s wage.\textsuperscript{26}

### 3.2.2 Value Functions

We derive and discuss the worker and firm value functions.

The process that governs the change in beliefs is a diffusion without a drift, so using Ito’s lemma we can write the flow value of an employed worker in occupation \( i \in \{W, B\} \) as:

\[
 rV_i(p) = w_i(p) + \frac{1}{2} \zeta_i^2 p^2 (1-p)^2 V''_i(p) \\
-\delta_i [V_i(p) - U(p)] - \gamma V_i(p) + \max \{ \eta_i \lambda (V_k(p) - V_i(p)), 0 \} \tag{3}
\]

where \( r \) denotes the worker’s and the firm’s discount rate, \( V_i(\cdot) \) is the value of the worker in occupation \( i \), \( w_i(\cdot) \) is the occupation-specific wage function and \( U(\cdot) \) 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, \( \zeta_i \). In our setup, in contrast to Jovanovic (1979), 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’s unemployed or employed in another occupation \( k \neq i \). The worker exogenously loses his job and becomes unemployed at rate \( \delta_i \), in which case his value is reduced to \( U(p) \). He dies at rate \( \gamma \). Finally, at rate \( \eta_i \lambda \) he meets a firm in the other occupation, which he joins if it is profitable for him to do so. Even though searching on-the-job is costless, it may still 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.

Similarly, the flow value to the firm of a filled vacancy in occupation \( i \) is given by:

\[
 rJ_i(p) = \bar{\alpha}_i(p) - w_i(p) + \frac{1}{2} \zeta_i^2 p^2 (1-p)^2 J''_i(p) \\
-\delta_i J_i(p) - \gamma J_i(p) - \eta_i \lambda J_i(p) I \{ V_i(p) < V_k(p) \} \tag{4}
\]

\textsuperscript{26}In Appendix A we also show that the linear sharing rule is bilaterally efficient in this case.
where \( J_i(\cdot) \) is the asset value of the firm, \( \pi_i(p) = p a_i w + (1 - p) a_i b \) 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’re assuming that the value of an unfilled vacancy is zero; in the online appendix we allow for free firm entry in each occupation, subject to an entry cost, which implies zero vacancy value.

Finally, the worker’s flow value of being unemployed is:

\[
 r U(p) = b_u + \max_i (V_i(p) - U(p)) - \gamma U(p) \tag{5}
\]

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, \( \lambda \), 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, unlike Jovanovic (1979) where beliefs are reset upon separation.

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

### 3.2.3 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 in finding a job in the other occupation.

Formally, the optimal decision rules are characterized by three threshold values of \( p \): \( p < \phi < \bar{p} \). In particular, \( \phi \) determines the level of the posterior at which a worker employed in occupation \( W \) optimally quits to unemployment. Similarly, a worker employed in occupation \( B \), optimally quits to unemployment when his posterior hits \( \bar{p} \). Finally, \( \phi \)
determines the optimal search decision for both unemployed and employed workers: if a worker is unemployed, he searches for employment in occupation $W$ if $p > \hat{p}$ and in occupation $B$ if $p \leq \hat{p}$; a worker employed in occupation $W$, searches on-the-job if $p \leq \hat{p}$; similarly, a worker employed in occupation $B$, searches on-the-job if $p > \hat{p}$.

Consider a worker employed in occupation $W$. As shown in appendix B, $V_W(p) - U(p)$ is increasing in $p$. Here $\bar{p}$ denotes the value of the posterior such that:

$$V_W(p) = U(p)$$

(or equivalently $J_W(p) = 0$). When $p$ reaches $\bar{p}$ the worker optimally quits to unemployment. Similarly for a worker employed in occupation $B$, we show that $V_B(p) - U(p)$ is decreasing in $p$. Here $\overline{p}$ denotes the value of the posterior such that:

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

(or equivalently $J_B(\overline{p}) = 0$). The worker optimally quits to unemployment when his $p$ reaches $\overline{p}$.

These thresholds can be seen in Figure 1, which displays the worker’s value functions in an economy in which the two occupations are symmetric. The worker’s value of being employed in occupation $W$ is always above his value of being unemployed, as long as $p > \bar{p}$. As $p$ converges to $\bar{p}$ the difference between the two values shrinks, until they become tangent (smooth pasting condition), at which point the worker quits and enters unemployment to search for a job in occupation $B$.

Finally, $\hat{p}$ denotes the value of the posterior at which the worker is indifferent between
searching for a job in occupation $W$ or $B$:

$$V_W (\hat{p}) = V_B (\hat{p})$$

For $p \geq \hat{p}$, the worker searches for employment in occupation $W$, while for $p < \hat{p}$, he searches in occupation $B$. Note that $\hat{p}$ also determines whether an employed worker searches on the job or not: as beliefs cross $\hat{p}$, the value of being employed in one occupation compared to the other changes. In appendix B we show that $\hat{p}$ is unique.

The cutoffs $\underline{p}$, $\hat{p}$ and $\overline{p}$ are determined by the solution of the system of equations (18)-(22) in appendix B. In appendix B, we also solve for the wage function, as well as the value of the firm, $J_i (\cdot)$ 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 or not. 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. In turns out that when the worker searches on the job, he compensates his firm by an amount equal to the weighted average of the worker’s gains, (in this case $V_k (p) - V_i (p)$), and the firm’s losses, $J_i (p)$, multiplied by his job finding probability. This generates a discontinuity in the wage function.

Finally, in Appendix C we derive analytically the stationary distribution of workers by expected productivity. The distribution of workers’ posteriors in occupation $W$ ($B$) features a fat right (left) Pareto-type tail, if and only if $\gamma > \frac{\zeta_B^2}{\delta_W + \lambda + \gamma}$ or approximately $\gamma > \zeta_W^2$ ($\gamma > \zeta_B^2$). These conditions have an intuitive economic interpretation: since 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, we derive conditions under which the within occupation cross-sectional distribution of wages features a fat, Pareto-type tail.

4 Model Implications

In this section we discuss the model’s implications regarding occupational choice, wage growth and occupational mobility, as well as examine the possible productivity parameter combinations that arise in our setup and how we can potentially differentiate between them.

We first examine occupational choice in this framework. In the absence of information frictions, workers sort according to their absolute advantage, without taking the other
type’s productivity into account. If there was only one occupation, learning would concern each individual worker’s ability, as in Harris and Holmstrom (1982). In contrast, in a world with information frictions and two (or more) occupations, workers base their occupational choice on the relative productivity differences between the two types in each occupation.

In particular, in a world with information frictions, three cases of interest may arise in the two-types, two-occupations framework. In the first, each worker-type is better than the other in his preferred occupation, i.e. a white type, who by assumption is better in $W$ than in $B$ ($a^w_W > a^w_B$), is also better than the blue type in $W$ ($a^w_W > a^b_W$); similarly, a blue type, who is most productive in $B$ ($a^b_B > a^b_W$), is also better than a white type in $B$ ($a^b_B > a^w_B$). This case is depicted in Figure 2, where we plot 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.

In the second case, one type, e.g. the white type, has absolute advantage in both occupations ($a^w_W > a^b_W$ and $a^w_B > a^b_B$), so now, as depicted in Figure 3, 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 depicted in Figure 4, the two types are equally productive in occupation $B$, so there is no learning in that occupation and occupation $B$ is an absorbing state. In this case, as in Jovanovic (1979), the worker learns only about the quality of his match in occupation $W$, while what he learns is not informative about his productivity
in occupation $B$, which is known.

In order to distinguish between the abovementioned cases, we examine our setup’s implications regarding wage growth—which in turn depends on the evolution of beliefs—and the within-occupation wage distributions. Market production contains information about a worker’s type, which is used to update beliefs, gradually revealing his true type. Although a worker’s posterior belief is a martingale, if we condition on his true type, his posterior is either strictly increasing or decreasing in expectation and the true type eventually becomes known. For example, for the case of a white type in occupation $i \in \{W, B\}$:

$$E( p_{t+\Delta t} \mid \text{true type is } w) = p_t + p_t (1 - p_t)^2 \zeta^2 \Delta t > p_t = E(p_t)$$

In other words, if his true type is $w$, his posterior converges in expectation to one. Similarly, if his true type is $b$ his posterior converges in expectation to zero.

When no worker-type has an absolute advantage, as in Figure 2, learning implies that the worker’s expected output increases over a long enough time horizon: white type workers eventually self-select to occupation $W$, while blue type workers choose occupation $B$; their prior converges to one and zero respectively, increasing their expected output. This is also reflected in their wage: the value of the match’s surplus is equal to the value of expected output plus the value of learning. As expected output increases, so does the surplus and since workers receive a constant share of the surplus as wage compensation, in this case our setup generates positive returns to labor market experience. Note however that it is not always the case that the wage is universally increasing in expectation, even
if the worker is in the “correct” occupation. In particular, it is possible that for some parameter values, as \( p \) approaches 0 or 1, the decline in the value of learning, is larger than the increase in expected output.\(^{27}\) In these cases, the wage declines before it picks up again, as the wage can’t be universally declining.\(^{28}\)

When, however, one worker-type is more productive in both occupations, as in Figure 3, expected output increases only for the white workers in the long-run. In this case, only the “high ability” workers enjoy positive returns to experience, while “low ability” workers see their wage decline over time, 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.

One could also differentiate between the two cases by looking at the support of the within-occupation wage distribution. In the case of “high ability”/“low ability” workers, our setup predicts very little overlap in the wage distributions as workers who are believed to be “high ability”, work in the “high ability” occupation and earn higher wages than workers in the “low ability” occupation. This should be even more pronounced when looking at younger workers who have yet to accumulate human capital. We discuss this further in Section 5.4 where we analyze the results of the empirical exercise.

Note also that if search frictions are not too large, we can use information on the location of workers who switch in the within-occupation wage distribution to distinguish 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, and they are also 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.

\(^{27}\)Note that the wage for occupation \( B \), when the worker is not searching on-the-job, is given by:

\[
\begin{align*}
    w_B (p) &= q\bar{a}_B (p) + q\lambda J_B (p) + (1 - q) b_u - \frac{q\lambda}{2(\tau + \gamma)} \zeta_{B}^{2} p^{2} (1 - p)^{2} J''_B (p)
\end{align*}
\]

We know that \( J''_B (p) < 0 \) and \( J''_B (p) > 0 \) (see Appendix B). As \( p \) approaches 0, both \( \bar{a}_B (p) \) and \( J_B (p) \) increase. However it’s possible that for certain parameter values, when \( p > 0.5 \), the decline in the last term is larger than the increase in the first two terms, leading to an overall decline in the wage.

\(^{28}\)To see this, note that for instance when \( p \) approaches 0 and all terms of order higher than 1 are dropped, the wage is given by:

\[
\begin{align*}
    w_B (p) &\simeq q\bar{a}_B (p) + (1 - q) b_u + \frac{q\lambda (1 - q) (\bar{a}_B (p) - b_u)}{\tau + \gamma + \delta_B + q\lambda}
\end{align*}
\]

which is decreasing in \( p \), if \( \bar{a}_B (p) \) is decreasing in \( p \). In other words, as \( p \) approaches 0, the wage increases.
mean of their new occupation.

Next, regarding returns to experience and wage dispersion, note that in our economy they are not the result of heterogeneity across workers in their observed ability levels (e.g. due to human capital accumulation), nor are they the result of search frictions (present in our setup), as in Burdett and Mortensen (1998). Returns to experience and wage dispersion here capture learning, as well as the selection of workers into the occupations best suited for them. In our setup, different workers are paid differently, both due to differences in their true abilities (types), as well as differences in the informativeness of the signals they have received. Furthermore, learning causes wages to increase because i) it allows workers to self-select into the occupation they perform best, and ii) it revises upwards their expected output in that occupation.

Moreover, consistent with empirical evidence, our framework can generate the decline in occupational mobility with experience: as workers eventually learn their type and self-select in the occupations they fit best, occupational mobility converges to zero as
the posterior converges to either zero or one.\textsuperscript{29} In addition, as one can observe in the empirical occupation transition matrix, Table 1, transitions from a given occupation, to certain occupations are more likely than to others, i.e. there exist regularities among the occupational transition patterns.\textsuperscript{30} Indeed in our setup, when there are more than two occupations, workers make occupational choices based on their past experience and therefore occupational choice is not random. Furthermore, as discussed in the previous section, our setup can replicate the fat Pareto tails of the within-occupation wage distributions.

The setup developed thus far considers the binary case with two possible states of the world; we relax this assumption in the next section where we perform a quantitative assessment of our setup.

\section{Quantitative Analysis}

In this section we perform a quantitative analysis with three objectives in mind. First, we want to assess the extent to which a relatively simple model can replicate the key features of the data, such as the observed occupational choices, the magnitude and the decline in occupational mobility, the observed wage distributions. Second, we wish to obtain estimates of some key parameters of the model and assess their quantitative importance, such as the speed of learning or the relative productivity distribution parameters. In particular, we want to understand whether sorting across the occupational groups considered takes place according to a hierarchical ability model, and thus can be modeled using a unidimensional ability measure, or according to comparative advantages. Third, we use the recovered parameters to consider the effect of an increase in the unemployment benefits on the economy’s productivity and calculate their welfare-maximizing level.

We begin this section by providing a generalization of our model which adds certain realistic features to our current framework. We then describe the data employed. We next turn to the calibration procedure and the results, while by discussing policy implications.

\textsuperscript{29}This implies that the separation rate to unemployment falls with age, as older workers are less likely to separate endogenously. Furthermore, extending the framework to allow workers to receive offers from both occupations and choose whether to accept them, would reproduce a declining job finding rate with age: younger workers are willing to work in both occupations and therefore accept all offers, whereas older ones can only form a profitable match in the occupation where they are expected to perform best.

\textsuperscript{30}The occupational transition matrix also sheds light onto the stepping-stone mobility model of Jovanovic and Nyarko (1997) who claim that some occupations serve as a "stepping-stone" to better, higher paid occupations (see also Sicherman and Galor (1990)). Both McCall (1990) and Kambourov and Manovskii (2008) construct similar occupation transition matrices and note that there are no asymmetries in the matrix that would lend support to the stepping-stone hypothesis.
5.1 General Model

Our setup in the previous section is tractable enough to deliver closed-form solutions and derive qualitative implications. Nonetheless, before taking our model to the data, we extend it to allow for features of the data which we know play an important role, such as human capital accumulation. Moreover, we allow for more than two worker types, in order to avoid having the entire wage dispersion being picked up by two points.

In particular, we extend our setup in the following dimensions: a) we allow for three instead of two occupations, b) we relax the binary restriction by assuming that the workers’ productivity (type) is drawn from a trivariate normal distribution 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 continuous time. We next specify the timing of events.

First consider a worker beginning the period unemployed:

a. He receives $b_u$ this period.

b. With probability $\lambda$, next period he’s employed in occupation $j$, while with probability $1 - \lambda$, he remains unemployed.

c. The death shock $\gamma$ is realized.

Next 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 an 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. The death shock $\gamma$ and job loss shocks, $\delta_i$ are realized.
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 chooses unemployment, he simply follows the timing of an unemployed worker.

We next 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 \( v_{1}^{k} \) about each occupation \( i \):

\[
\begin{align*}
    v_{1}^{k} & \sim N \left( \mu_{1}, \kappa_{1}^{2} \right) \\
v_{2}^{k} & \sim N \left( \mu_{2}, \kappa_{2}^{2} \right) \\
v_{3}^{k} & \sim N \left( \mu_{3}, \kappa_{3}^{2} \right)
\end{align*}
\]

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

\[
m^{k} = \begin{bmatrix}
m_{1}^{k} \\
m_{2}^{k} \\
m_{3}^{k}
\end{bmatrix} \sim N \left( \begin{bmatrix}
v_{1}^{k} \\
v_{2}^{k} \\
v_{3}^{k}
\end{bmatrix}, 
\begin{bmatrix}
\tau_{1}^{2} & \rho_{12}\tau_{1}\tau_{2} & \rho_{13}\tau_{1}\tau_{3} \\
\rho_{12}\tau_{1}\tau_{2} & \tau_{2}^{2} & \rho_{23}\tau_{2}\tau_{3} \\
\rho_{13}\tau_{1}\tau_{3} & \rho_{23}\tau_{2}\tau_{3} & \tau_{3}^{2}
\end{bmatrix} \right)
\]

and is unobserved.

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

The match output produced from a match \( \kappa \) between worker \( k \) and a firm in occupation \( i \) is given by:

\[
y_{i}^{\kappa} = m_{i}^{k} + g_{HC} \left( x \right) + \sigma_{i}\varepsilon_{i}^{\kappa}
\]

where \( \varepsilon_{i}^{\kappa} \sim N \left( 0, 1 \right) \).

As in our benchmark model, a worker observes his match output and updates his beliefs about his underlying type using Bayes’ rule. His beliefs regarding his type \( m^{k} \) also follow a trivariate normal distribution with mean \( \mu \) and variance \( \Sigma \). In Section 4.1 of the online appendix, we describe the updating of beliefs.

Before presenting the value functions, it is useful to define the value \( C \left( \mu, \Sigma, x \right) \), which captures the value of a worker with beliefs \( \left( \mu, \Sigma \right) \) and experience \( x \), who this period has
the option of becoming employed in any occupation:

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

where \(\beta\) is the discount factor, \(V_i(\cdot)\) is the value of an employed worker in occupation \(i\) and \(U(\cdot)\) is the value of an unemployed worker. Finally \(\Sigma'(i)\) represents the value of the updated beliefs’ variance-covariance matrix which, as described in Section 4.1 of the online appendix, is updated deterministically, but depends on the occupation \(i\) the worker has been employed in. 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.

The value function of a worker currently employed in occupation \(i\) is:

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

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 searches on-the-job and finds a job in this preferred occupation (\(\arg \max C(\mu, \Sigma, x)\)) with probability \(\eta_i \lambda\) and if not, he continues working in occupation \(i\), or he immediately quits to unemployment. Note that the wage function differs depending on whether the worker is searching on the job or not (\(w_i^{NS}(\cdot)\) vs. \(w_i^{OTJS}(\cdot)\)).

The value of an unemployed worker is given by:

\[
U(\mu, \Sigma, x) = b_u + \beta (1 - \gamma) \lambda C(\mu, \Sigma, x) + \beta (1 - \gamma) (1 - \lambda) U(\mu, \Sigma, x)
\]

The worker receives \(b_u\) and with probability \(\lambda\) he begins employment next period 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:

$$qJ_i (\mu, \Sigma, x) = (1 - q) [V_i (\mu, \Sigma, x) - U (\mu, \Sigma, x)] \quad (8)$$

In the online appendix we show that the wage function in each case is given by:

$$w_{i}^{NS} (\mu, \Sigma, x) = q (\mu_i + g_{HC} (x)) + (1 - q) U (\mu, \Sigma, x) - \beta (1 - \gamma) (1 - q) E \mu U (\mu', \Sigma' (i), x')$$

$$w_{i}^{OTJS} (\mu, \Sigma, x) = q (\mu_i + g_{HC} (x)) + \frac{1 - q}{1 - \eta_i \lambda} U (\mu, \Sigma, x) - \beta (1 - \gamma) (1 - q) E \mu U (\mu', \Sigma' (i), x') - \frac{\eta_i \lambda}{1 - \eta_i \lambda} (1 - q) C \mu (\mu, \Sigma, x)$$

The solution to the worker’s problem i.e. determine his preferred occupation, whether to search on-the-job and when to quit to unemployment, has to be derived numerically.

The above setup is more general than our benchmark specification, but it does not admit closed form solutions and thus its implications can only be studied quantitatively.

5.2 Data

The information necessary for our quantitative exercise includes workers’ wages, occupational affiliation and employment status, as well as worker transitions between occupations and to/from unemployment. The 1996 panel of the SIPP suits our purposes well. It is designed to be a nationally representative sample of households in the civilian non-institutionalized U.S. population, with interviews being conducted every four months for four years. Each interview records information about the worker’s current occupation and wage, as well as a complete weekly employment history for the past four months. 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 can also explain differences in wages (Abowd et al. (1999)) and potentially in occupational mobility as well.
(Kambourov-Manovskii (2008)). We therefore, use only subsamples of workers that share the same observable characteristics: white males with a high school degree. In Section 8 of the online appendix we present results for white males with some college education, as well as those with a completed college education. We further restrict our sample by using workers up to the age of 45, as we expect that 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 individuals who are enrolled in school for the duration of their studies, as well as for the period before them. Finally, we do not use observations for which labor force data have been imputed, nor any observations for which the occupation may have been potentially miscoded.\textsuperscript{31}

Our occupational partition for high school graduates focuses 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 laborers, such as various machine operators (occupation 3).\textsuperscript{32} Using a finer partition is computationally prohibitive. The 1996 panel of the SIPP uses dependent interviewing, which is found to reduce occupational coding error (Hill (1994)). The partition of the occupational codes into these three categories can be found in Section 5 of the online appendix.

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 real 1996 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).

5.3 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

\textsuperscript{31}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.

\textsuperscript{32}For workers with some college education and completed college education, we focus on white-collar, blue collar and pink-collar jobs as in Lee and Wolpin (2006).
20 years of working experience and we recover the true productivity distributions using the occupational choices and wages of older workers.\textsuperscript{33} 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 corresponding to the interval between interview waves in our data and the discount rate, $\beta$, is set to 0.9901, which implies an annual rate of approximately 3%. As we do not have information on firm profits, we calibrate the value of the Nash bargaining coefficient, $q$. Since in our framework we allow for on-the-job search, we choose $q$ equal to 0.3, which is on the high end of the estimates in Cahuc et al. (2006), but lower than the estimates of models with no on-the-job search, such as Flinn (2006). We set the death rate parameter, $\gamma$, to 0.013333 which implies an average working life of approximately 25 years.\textsuperscript{34} 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 rate $\lambda$, which we calibrate following Shimer (2012) - see Section 6 of the online appendix for details. The exogenous separation rates, $\delta_i$, 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. Since the SIPP contains weekly employment information, time aggregation is not a big concern when calculating these transition moments. Table 2 shows the resulting parameters.

The remaining parameters are calibrated in two steps, which are performed iteratively. In the first step we recover the parameters governing the true productivity distribution. More specifically, in the online appendix we show that the productivity parameters follow

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
$\delta_1$ & 0.017 & $\lambda$ & 0.828 \\
$\delta_2$ & 0.018 & $b_u$ & $\$3.33$ \\
$\delta_3$ & 0.016 & \\
\hline
\end{tabular}
\end{center}

Table 2: Parameters Calibrated Independently

\textsuperscript{33}Using a different approach, Lange (2006) finds that employer learning about a worker’s type is relatively fast.

\textsuperscript{34}Using a smaller value for $\gamma$ would imply that the average age in the cross-section of workers is unrealistically large.
the following trivariate normal:

\[
\begin{bmatrix}
    m^k_1 \\
    m^k_2 \\
    m^k_3 
\end{bmatrix} \sim N
\begin{pmatrix}
    \mu_1 \\
    \mu_2 \\
    \mu_3 
\end{pmatrix},
\begin{pmatrix}
    \kappa^2_1 + \tau^2_1 & \rho_{12}\tau_1\tau_2 & \rho_{13}\tau_1\tau_3 \\
    \rho_{12}\tau_1\tau_2 & \kappa^2_2 + \tau^2_2 & \rho_{23}\tau_2\tau_3 \\
    \rho_{13}\tau_1\tau_3 & \rho_{23}\tau_2\tau_3 & \kappa^2_3 + \tau^2_3 
\end{pmatrix}
\] (9)

The value and wage of older workers, who are assumed to have learned their type, is given by (15A) and (16A) respectively of the online 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, \(\delta_i\). 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, eq.(9), conditional on the human capital accumulation process, \(g_{HC}(x)\), i.e.

\[
\phi_1 = \{\mu_1, \mu_2, \mu_3, \kappa^2_1 + \tau^2_1, \kappa^2_2 + \tau^2_2, \kappa^2_3 + \tau^2_3, \rho_{12}\tau_1\tau_2, \rho_{13}\tau_1\tau_3, \rho_{23}\tau_2\tau_3\}
\]

In order to recover parameters \(\phi_1\), 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 3 occupations, variance and skewness of the wage distribution in each of the 3 occupations and share of wages over $22 in each occupation.\(^{35}\) 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 next turn to the remaining parameters we need to recover:

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

i.e. the parameters associated with the human capital accumulation process, \(g_{HC}(x)\), \(\omega_1\) and \(\omega_2\) -see below-, the standard deviation of output realizations, \(\sigma_i\), the standard deviation of the initial beliefs \(\kappa_i\) and the parameters governing the effectiveness of on-the-job search, \(\eta_i\). Note that from step 1, we have pinned down \(\kappa^2_i + \tau^2_i\) and \(\rho_{ij}\tau_i\tau_j\), so that by choosing \(\kappa_i\) in step 2, we also determine \(\tau_i\) and \(\rho_{ij}\).

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

\(^{35}\)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 discussion in French and Taber (2011)).
of experience:

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

We recover the parameters \( \phi_2 \) by matching the following moments with their empirical counterparts: for workers with less than 2 years of labor market 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 less than 4 years of labor market experience, as well as for workers with 4 to 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. Since the choice of the human capital parameters affects the first stage results, the two stages are performed iteratively. Section 7 of the online appendix contains details of the procedure.

Although a rigorous identification argument is impossible due to the complexity of our framework, we attempt to give an informal argument of 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, \( \kappa_i \). The rate at which workers switch occupations as well as its decrease specify the speed at which workers learn about their underlying productivities, which is governed by \( \sigma_i \). The fraction of job-to-job transitions pins down the efficiency of on-the-job search, \( \eta_i \). 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 \( \sigma_i \), we can pin down the rate of human capital accumulation.

### 5.4 Results

The resulting parameters are reported in Table 3, while the moments matched are reported in Table 4.

Our results suggest that, on average, workers are less productive in occupation 2 than in the other two occupations, since \( \mu_2 < \mu_1, \mu_3 \). However the standard deviation of the productivity draw, \( \tau_2 \), is larger for that occupation. The high standard deviation of productivity explains the large fraction of workers that eventually chose occupation 2 (37.56% as shown in Table 4). The low dispersion of initial beliefs (\( \kappa_i \)) leads to a lower
share in occupation 2 (26.82%) among younger workers, as these workers initially believe they may be better suited in one of the other 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 (τ^{k}_i)*, 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 $3.09.36

The human capital profile is initially upward slopping and then downward slopping, potentially indicating human capital decumulation. Bagger et al. (2011) also report a gradual loss of productivity that is more pronounced for less educated workers.

The results also suggest that the allocation of workers across these occupations is not well described by the hierarchical model of ability in which some workers perform better in all occupations. Even though, our setup allows for a hierarchical ranking of occupations, we find that this is not supported by the data (for instance the correlation coefficient ρ_{23} is equal to -0.934). These results are related to work by Groes et al. (2010) who present evidence in favor of vertical sorting across occupations. Our empirical exercise focuses on three broad occupational groups, whereas they look at finer classifications. Our findings are not surprising. For expositional purposes, consider the case of two occupations: the hierarchical ability model implies that there is little overlap in the wage distributions of these two occupations, since workers who are thought to be the high ability ones, are employed in the high ability occupation and earn higher wages. In contrast, workers suspected of being low ability all work in the same occupation. In the extreme case, with no search frictions and no differences in the speed of learning, every worker in the low-

<table>
<thead>
<tr>
<th>μ_1</th>
<th>7.08</th>
<th>ω_2</th>
<th>-0.28</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ_2</td>
<td>6.09</td>
<td>ρ_{12}</td>
<td>0.011</td>
</tr>
<tr>
<td>μ_3</td>
<td>7.01</td>
<td>ρ_{13}</td>
<td>-0.161</td>
</tr>
<tr>
<td>τ_1</td>
<td>4.02</td>
<td>ρ_{23}</td>
<td>-0.934</td>
</tr>
<tr>
<td>τ_2</td>
<td>10.35</td>
<td>σ_1</td>
<td>12.01</td>
</tr>
<tr>
<td>τ_3</td>
<td>8.18</td>
<td>σ_2</td>
<td>14.44</td>
</tr>
<tr>
<td>κ_1</td>
<td>3.66</td>
<td>σ_3</td>
<td>9.99</td>
</tr>
<tr>
<td>κ_2</td>
<td>1.03</td>
<td>η_1</td>
<td>0.35</td>
</tr>
<tr>
<td>κ_3</td>
<td>1.02</td>
<td>η_2</td>
<td>0.204</td>
</tr>
<tr>
<td>ω_1</td>
<td>0.157</td>
<td>η_3</td>
<td>0.199</td>
</tr>
</tbody>
</table>

Table 3: Calibration Results

36See updating equations in Section 4.1 of the Online Appendix.
Table 4: Empirical and Predicted Moments. "Young" refers to workers between the age of 18 and 20, while "Old" refers to workers with more than 20 years of potential labor market experience (up to the age of 45). Occupational switching rates are 8-month rates. They are computed during the first and second 4 year periods of labor market experience.

Indeed, the wage distributions of the three occupations do not differ much (see moments in Table 4) and as expected, our results do not favor the hierarchical model of ability. Willis and Rosen (1979) also find evidence against the hierarchical ability model in a very different context, the decision to attend college. The results presented here are complementary to those in Groes et al. (2010), since taken together they 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
<table>
<thead>
<tr>
<th>Moment</th>
<th>Baseline</th>
<th>No Learning</th>
<th>No Search Frictions</th>
<th>No OTJS</th>
<th>No Initial Disp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young % occup 1</td>
<td>20.51%</td>
<td>38.14%</td>
<td>50.75%</td>
<td>31.52%</td>
<td>13.65%</td>
</tr>
<tr>
<td>Young % occup 2</td>
<td>34.19%</td>
<td>2.01%</td>
<td>33.02%</td>
<td>46.54%</td>
<td>19.77%</td>
</tr>
<tr>
<td>Young Mean Wage occ 1</td>
<td>$8.16</td>
<td>$9.06</td>
<td>$6.67</td>
<td>$7.20</td>
<td>$10.96</td>
</tr>
<tr>
<td>Young Mean Wage occ 2</td>
<td>$4.85</td>
<td>$6.45</td>
<td>$9.99</td>
<td>$3.22</td>
<td>$6.97</td>
</tr>
<tr>
<td>Young Mean Wage occ 3</td>
<td>$5.35</td>
<td>$6.57</td>
<td>$12.78</td>
<td>$6.57</td>
<td>$5.45</td>
</tr>
<tr>
<td>Young St Dev Wage 1</td>
<td>$4.21</td>
<td>$0.95</td>
<td>$3.37</td>
<td>$3.76</td>
<td>$2.94</td>
</tr>
<tr>
<td>Young St Dev Wage 2</td>
<td>$8.04</td>
<td>$1.60</td>
<td>$4.98</td>
<td>$9.89</td>
<td>$6.82</td>
</tr>
<tr>
<td>Young St Dev Wage 3</td>
<td>$7.07</td>
<td>$0.58</td>
<td>$4.30</td>
<td>$7.34</td>
<td>$6.67</td>
</tr>
<tr>
<td>Young Skewness 1</td>
<td>-0.24</td>
<td>0.48</td>
<td>-0.08</td>
<td>-0.02</td>
<td>-2.05</td>
</tr>
<tr>
<td>Young Skewness 2</td>
<td>0.57</td>
<td>1.03</td>
<td>1.32</td>
<td>0.68</td>
<td>0.21</td>
</tr>
<tr>
<td>Young Skewness 3</td>
<td>0.21</td>
<td>2.79</td>
<td>0.61</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>Occ Sw 1 (0-4 years)</td>
<td>14.84%</td>
<td>0</td>
<td>50.72%</td>
<td>4.79%</td>
<td>9.22%</td>
</tr>
<tr>
<td>Occ Sw 2 (0-4 years)</td>
<td>15.14%</td>
<td>0</td>
<td>31.38%</td>
<td>14.92%</td>
<td>10.33%</td>
</tr>
<tr>
<td>Occ Sw 3 (0-4 years)</td>
<td>14.72%</td>
<td>0</td>
<td>18.01%</td>
<td>7.14%</td>
<td>12.88%</td>
</tr>
<tr>
<td>Occ Sw 1 (5-8 years)</td>
<td>6.51%</td>
<td>0</td>
<td>11.85%</td>
<td>1.44%</td>
<td>5.88%</td>
</tr>
<tr>
<td>Occ Sw 2 (5-8 years)</td>
<td>2.92%</td>
<td>0</td>
<td>7.10%</td>
<td>1.16%</td>
<td>3.80%</td>
</tr>
<tr>
<td>Occ Sw 3 (5-8 years)</td>
<td>2.58%</td>
<td>0</td>
<td>6.26%</td>
<td>1.27%</td>
<td>2.37%</td>
</tr>
<tr>
<td>% Job-to-Job Sw occup 1</td>
<td>92.08%</td>
<td>NaN</td>
<td>1</td>
<td>0</td>
<td>77.79%</td>
</tr>
<tr>
<td>% Job-to-Job Sw occup 2</td>
<td>49.98%</td>
<td>NaN</td>
<td>1</td>
<td>0</td>
<td>76.33%</td>
</tr>
<tr>
<td>% Job-to-Job Sw occup 3</td>
<td>58.52%</td>
<td>NaN</td>
<td>1</td>
<td>0</td>
<td>56.32%</td>
</tr>
<tr>
<td>Mean Wage 30</td>
<td>$12.06</td>
<td>$8.72</td>
<td>$18.29</td>
<td>$11.97</td>
<td>$11.73</td>
</tr>
</tbody>
</table>

Table 5: Counterfactuals

above results.

Table 4 reports the predicted and actual moments. Our setup matches well the shares of older workers across occupations, as well as the wage means and standard deviations. It slightly underpredicts the degree of right skewness while it matches the right tail for workers in occupations 2 and 3 (those making more than $22), but underpredicts the right tail for workers in occupation 1. It does a reasonably good job of matching the shares of younger workers, but underpredicts the mean wage of young workers in occupation 1 and overpredicts those in occupations 2 and 3. It fails to match the (low) dispersion of young workers’ wages\(^{37}\), but mostly matches the level and decline of occupational mobility. It somewhat overpredicts the degree of job-to-job mobility for occupation 1 and the mean wage level at age 30.

Moreover, as discussed earlier, we can explore the wage of workers who switch occupations relative to other workers in the same occupation, both in the data, as well as in

\(^{37}\)The calibration effectively uses the \(\kappa_i\) parameters to match the shares, the mean wages, the standard deviation and the skewness of young workers’ wages, so these three parameters are overidentified.
our calibrated model. In the data, the average pre-switch wage of workers is lower than
the mean in their corresponding occupation, while 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 is also consistent with both facts in all 3 occupations, although the
differences between switchers and stayers are somewhat larger.\footnote{When controlling
for a fourth-order age polynomial, 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 occupa-
tion 2, $-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. Furthermore
in the data the residual post-switch mean wage of switchers in occupation 1 are $-0.97 versus $-1.01 for
stayers, $0.07 versus $1.18 for occupation 2, $-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.} As discussed in Section
4, this is also consistent with the calibration’s result that the hierarchical model of ability
in which some workers perform better in all occupations, does not describe well the alloca-
tion of workers across these occupations. It is also worth noting that there is a significant
number of switches involving a wage cut (35\%) in the calibrated model, somewhat more
than in the data (19\%).\footnote{This reflects both the option value of learning one’s ability in
another occupation, but also 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’s immediately successful, then the low realization does not show up
as a low wage in his original occupation in the following period.} Moreover, the setup generates significant return mobility, as
31\% of switchers eventually return to their original occupation. Finally looking at the
wage growth of workers who remain in the same occupation between the ages of 20 and
24, we observe that in the model their wages on average grow by 66.74\% in occupation
1 compared to 54.87\% in the data, 32.37\% in occupation 2, compared to 34.25\% in the
data and 13.22\% in occupation 3, compared to 46.69\% in the data.

To better understand the impact of the various elements of the model, we perform
a number of counterfactual exercises. First, we shut down learning and assume workers
keep their prior forever. Since there is no learning, workers remain in the same occupa-
tion forever. This experiment provides insights regarding the importance of information
frictions and learning. Wage dispersion is now only due to differences in the priors and
human capital accumulation. As shown in the second column of Table 5 the share of
occupation 2 is reduced to 2\%: given the high value of $\tau_2$ compared to $\kappa_2$, workers who
had high draws in occupation 2 do not know this and choose not to work there, given
the low dispersion of priors. Moreover, wage growth is reduced significantly, as it is now
driven only by general human capital 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 next shut down search frictions and allow workers to contact potential employers frictionlessly. 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 5, wages increase, as does occupational mobility.

If we shut down the ability of workers to search while employed, occupational mobility falls substantially (fourth column of Table 5): imposing the extra cost of having to quit to unemployment in order to switch occupations, reduces significantly the number of realized occupational transitions, given that a large fraction of switches involve employed workers.

Finally, we eliminate the initial dispersion of beliefs, i.e. set $\kappa_i = 0$, for all $i$ in the last column of Table 5. Now, all workers start with the same prior and thus make the same occupational choice, preferring to be employed in occupation 3, which has a high mean, $\mu_3$, as well as a high $\tau_3$ and low noise, $\sigma_3$. 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 unlike the previous counterfactuals, also alters the true productivity distribution.

### 5.5 Policy Implications

We next use our model to investigate the impact of increased unemployment on labor allocation. In the 1970s, the US and several European countries experienced large increases in their unemployment rate 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 here, 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 talents. Workers are now more unsure about their abilities on average and therefore more likely to be employed in an occupation that does not match their talents.

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 US and Europe in the 1970s. More specifically, we permanently decrease the job finding rate, $\lambda$, in order 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 labor market differences between Europe and the US.

Table 6 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 is $13.81. When we generate an increase in the unemployment to approximately 10%, output per employed worker drops to $13.02 per hour. Following our previous discussion, as workers now spend more time unemployed, workers are on average less informed regarding their true type and, therefore, are mismatched more often.

The above result has important implications for policy makers: economic policies that affect search frictions influence the economy’s allocation of resources and therefore 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, leads to a higher separation rate, as the outside value of the worker increases and matches that were acceptable before, now yield a negative surplus. In addition, an increase in the $b_u$ decreases profits of filled vacancies and dampens incentives to post vacancies in the first place, and leads to lower job finding rates.

To 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 is 50% of the average wage, as opposed to 30%. The increase in $b_u$, alters incentives for firm entry. In order 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.

<table>
<thead>
<tr>
<th></th>
<th>$b_u = 3.33$</th>
<th>$b_u = 3.33$</th>
<th>$b_u = 5.55$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_1 = \lambda_2 = \lambda_3 = 0.83$</td>
<td>$\lambda_1 = \lambda_2 = \lambda_3 = 0.44$</td>
<td>$\lambda_1 = 0.7, \lambda_2 = 0.75, \lambda_3 = 0.71$</td>
</tr>
<tr>
<td>Unemployment</td>
<td>5.83%</td>
<td>10.07%</td>
<td>6.89%</td>
</tr>
<tr>
<td>Output per Worker</td>
<td>$13.81$</td>
<td>$13.02$</td>
<td>$13.55$</td>
</tr>
</tbody>
</table>

Table 6: Impact of Search Frictions on Productivity
As shown in the third column of Table 6, 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 in the new steady state.\(^{40}\) Moreover, if we calculate welfare in the two cases, assuming unemployment benefits are financed through lump-sum taxes on all workers, we find that welfare is lower in the case of the increased \( b_u \).\(^{41}\) 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 previous literature: both Acemoglu and Shimer (1999) and Marimon and Zilibotti (1999) have illustrated mechanisms under which such an increase leads to improved resource allocation.\(^{42}\) Here the increase in unemployment insurance leads to a higher unemployment rate and to a deterioration of the allocation of workers across productive activities. Although estimating a model that incorporates all the above mechanisms is beyond the scope of this paper, our results 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)).

### 6 Conclusion

This paper investigates the labor market implications of worker learning about their unobserved abilities. We argue that this type of learning is consistent with the observed patterns of occupational mobility: the large bidirectional flows across occupations, the decline in the occupational switching probability as workers grow older and the non-random mobility of workers across occupations. Moreover, as the calibration of our framework illustrates, occupations with similar wage distributions are inconsistent with the hierarchical ability setup in which some workers perform better in all occupations; on the contrary, a setup in which each worker performs best in some occupation fits the data better.

The setup developed here has focused on labor market learning and occupational mobility. However it may speak to other contexts as well. For instance, the large flows of workers across space (Coen-Pirani (2010)) are consistent with migrants learning about

\(^{40}\)Moreover the percent of workers who choose to remain out of the labor force, which is negligible in our baseline calibration, rises to 5.3% in the case of increased \( b_u \).

\(^{41}\)Welfare, as in Flinn (2006), is calculated by the weighed sum of the average welfare of employed workers, unemployed workers, workers out of the labor force and filled vacancies, with weights given by the size of each group.

\(^{42}\)In Acemoglu and Shimer (1999) increased unemployment insurance encourages risk averse workers to apply for higher quality jobs, which are more difficult to obtain. Marimon and Zilibotti (1999) show that increased benefits act as a search subsidy allowing workers to obtain more suitable jobs.
the quality of their match with a particular location. In turn, the underlying qualities of matches are likely not independently distributed across locations. This can be particularly important when considering the observed prevalence of return migration (Kennan and Walker (2011)).

Markets for experience goods constitute another salient example. Consider a consumer who is learning whether a particular product suits his needs; for instance, a patient trying out a drug. The existing literature has examined both the case of learning about a common component across consumers (Milgrom and Roberts (1986), Bergemann and Välimäki (1996), Bose et al. (2006 & 2008), Bonatti (2011)), as well as the case of idiosyncratic uncertainty, where each consumer learns about his own individual match with a product (Crémer (1984), Erdem and Keane (1996), Israel (2005), Bergemann and Välimäki (2006)). A common assumption is that the information the individual obtains when consuming one product is not useful in evaluating his match with a different product. Again, one might expect to find correlation in the match qualities across products so that the consumer experimenting with one product learns about his match with other firms’ products (e.g. how he reacts not just to this particular drug, but also to drugs with similar characteristics or chemical structure). He might also infer his match about other products of the same firm.\footnote{Gavazza (2011) argues that firms that offer a greater variety of products have an advantage over those that offer fewer products.}

In the same vein, one might consider the problem of a firm introducing more than one new products with correlated, but unknown, demands. For instance, a cereal company might introduce different varieties of a new product or a car manufacturer might introduce different designs or specifications of a new model. In this case, the information flow obtained from selling one product is informative about the popularity of the other current or future products. The optimal pricing decision of the firm inherits this interdependence, and induces an interesting trade-off between current profits, information about the popularity of the current product and information about other products.

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 lump-sum 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 backwards 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 Figure 5 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 \(\varepsilon\) 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, as he 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 and the subgame perfect equilibrium here is \((N-I_2, N-P)\), 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’ll lose the auction), in which case the worker moves to the poacher and receives no transfer. In the case where the incumbent declines to enter in the first stage, then the poacher foresees that he won’t enter in the second stage either and chooses to pay \(\varepsilon\) 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, so he chooses not enter in the first stage. Thus the subgame perfect equilibrium here is \((N-I_2, Y-P)\), 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. If 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 $\varepsilon$, 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 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 won’t enter, whereas if the incumbent does not enter in the first stage, he knows that the poacher will and he won’t enter in the second stage either. Thus the subgame perfect equilibrium is $(Y-I_1, N-P)$, 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).\footnote{Mortensen (2003) points out, “unlike in the market for academic economists in the United States, making counteroﬀers is not the norm in many labor 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 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, if 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 doesn’t 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 wouldn’t 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

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 $V (\cdot)$ is a convex function of beliefs. The same argument 
ensures that $J_i (\cdot)$ and $U (\cdot)$ are also convex. The convexity of $J_i (\cdot)$ implies that $\underline{p}$ and $\bar{p}$ 
are unique.
Furthermore, \( J_W (\cdot) \) is globally increasing in \( p \). To understand this, note that \( J_W (p) \geq 0 \ \forall p, \ \text{and} \ J_W (\bar{p}) = 0 \). Since \( J_W (\cdot) \) is convex, then it must be increasing everywhere. Intuitively, for \( \tilde{p} \) slightly larger than \( p \), \( J_W (\tilde{p}) > 0 \), and thus it must be increasing in that region. Convexity ensures that it is increasing everywhere. Similarly we can show that \( J_B (\cdot) \) is decreasing in \( p \).

To see that the threshold \( \bar{p} \) is unique, remember that it is defined from \( V_W (\bar{p}) - U (\bar{p}) = V_B (\bar{p}) - U (\bar{p}) \). However from the Nash bargaining solution we know that \( qJ_i (p) = (1 - q) [V_i (p) - U (p)] \Leftrightarrow V_i (p) - U (p) = \frac{q}{1 - q} J_i (p) \). Monotonicity and convexity of \( J_i (\cdot) \) 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 \( \lim_{p \to 0} V_B (p) > \lim_{p \to 0} V_W (p) \) and that \( \lim_{p \to 1} V_W (p) > \lim_{p \to 1} V_B (p) \) (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.

Let’s start with the case where the worker’s outside option is the occupation he is currently employed in.

Now (3) through (5) become:

\[
\begin{align*}
  rV_i (p) & = w_i (p) + \frac{1}{2} \zeta_i^2 p^2 (1 - p)^2 V_i'' (p) - \delta_i [V_i (p) - U (p)] - \gamma V_i (p) \\
  rJ_i (p) & = \bar{a}_i (p) - w_i (p) + \frac{1}{2} \zeta_i^2 p^2 (1 - p)^2 J_i'' (p) - \delta_i J_i (p) - \gamma J_i (p) \\
  rU (p) & = b_u + \lambda [V_i (p) - U (p)] - \gamma U (p)
\end{align*}
\]

We subtract the worker’s flow value of being unemployed (eq. (5')) from his flow value of being employed (eq. (3')) and multiply through by \( (1 - q) \):

\[
\begin{align*}
  (1 - q) r (V_i (p) - U (p)) & = (1 - q) (w_i (p) - b_u) + \frac{1}{2} (1 - q) \zeta_i^2 p^2 (1 - p)^2 V_i'' (p) \\
  & - \delta_i (1 - q) [V_i (p) - U (p)] - \gamma (1 - q) V_i (p) \\
  & - \lambda (1 - q) [V_i (p) - U (p)] + \gamma (1 - q) U (p)
\end{align*}
\]
We similarly multiply the flow asset value of a filled vacancy (eq. (4')) by $q$:

$$ qrJ_i(p) = q\bar{a}_i(p) - qw_i(p) + \frac{1}{2}q\zeta_i^2p^2(1-p)^2J''_i(p) $$

$$ -q\delta_iJ_i(p) - q\gamma J_i(p) $$

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

$$ w_i(p) - (1-q)b_u + \frac{1}{2}(1-q)\zeta_i^2p^2(1-p)^2V''_i(p) $$

$$ -\lambda(1-q)[V_i(p) - U(p)] - q\bar{a}_i(p) - \frac{1}{2}q\zeta_i^2p^2(1-p)^2J''_i(p) = 0 \Leftrightarrow $$

$$ w_i(p) = q\bar{a}_i(p) + (1-q)b_u + q\lambda J_i(p) - \frac{1}{2}(1-q)\zeta_i^2p^2(1-p)^2V''_i(p) $$

$$ + \frac{1}{2}q\zeta_i^2p^2(1-p)^2J''_i(p) \quad (10) $$

Using the surplus sharing condition once again:

$$ qJ''_i(p) = (1-q)(V''_i(p) - U''(p)) $$

$$ V''_i(p) = \frac{q}{1-q}J''_i(p) + U''(p) \quad (11) $$

However from the value of the unemployed worker (eq. (5')), we have:

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

$$ U''(p) = \frac{\lambda}{r + \gamma + \lambda}V''_i(p) $$

Substituting out for $U''(p)$ in (11) results in:

$$ V''_i(p) = \frac{q}{1-q}J''_i(p) + \frac{\lambda}{r + \lambda + \gamma}V''_i(p) \Rightarrow $$

$$ V''_i(p) = \frac{q}{1-q}\frac{r + \gamma + \lambda}{r + \gamma}J''_i(p) \quad (12) $$

We can now substitute out for $V''_i(p)$ in (10) and obtain the wage as a function of
where $\lambda$ is the 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 $q \left( J_i''(p) + V_i''(p) \right) - V_i''(p) < 0$, 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 $i \in \{W, B\}$ (eq. (3’)) we get:

$$rV_i(p) = (1 - q) \left( b_u + \lambda [V_i(p) - U(p)] \right) + q \left( \bar{a}_i(p) + \frac{1}{2} \zeta_i^2 p^2 (1 - p)^2 \right)$$

$$\left( J_i''(p) + V_i''(p) \right)$$

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

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) and 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 and 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 \( J_i(p) \) in the case in which the worker’s outside option is searching for another job in his current occupation:

\[
(r + \gamma + \delta_i + q \lambda) J_i(p) = (1 - q) (\bar{a}_i(p) - b_u) + \frac{r + \gamma + q \lambda}{2 (r + \gamma)} \zeta_i^2 p^2 (1 - p)^2 J''_i(p) (14)
\]

and in the case in which it is not:

\[
(r + \gamma + \delta_i + \eta_i \lambda) J_i(p) = (1 - q) (\bar{a}_i(p) - b_u) + \frac{q \lambda}{2 (r + \gamma)} \zeta_i^2 p^2 (1 - p)^2 J''_i(p) (15)
\]

\[+ \frac{1}{2} \zeta_i^2 p^2 (1 - p)^2 J''_i(p) - q \lambda (1 - \eta_i) J_k(p)\]

The general solution to differential equation (14) is:

\[J_i(p) = \frac{(1 - q) (\bar{a}_i(p) - b_u)}{r + \gamma + \delta_i + q \lambda} + \frac{K_1' p^{1/2} \frac{1}{2} \sqrt{\frac{4 + h_i}{h^i}}}{(1 - p)^{1/2} \sqrt{\frac{4 + h_i}{h^i}}}
\]

\[+ K_2' p^{1/2} \frac{1}{2} \sqrt{\frac{4 + h_i}{h^i}} (1 - p)^{3/2} \frac{1}{2} \sqrt{\frac{4 + h_i}{h^i}}\]

where \( h_i = \frac{1}{2} \frac{r + \gamma + q \lambda}{(r + \gamma + \delta_i + q \lambda) (r + \gamma + \delta_i + q \lambda) \zeta_i^2} \) and \( K_1' \) and \( K_2' \) are undetermined coefficients. For the case of \( i = W \), when \( p \to 1, \lim_{p \to 1} K_2^W p^{1/2} \frac{1}{2} \sqrt{\frac{4 + h_W}{h^W}} (1 - p)^{1/2} \frac{1}{2} \sqrt{\frac{4 + h_W}{h^W}} = K_2^W \cdot 1 \lim_{p \to 1} (1 - p)^{1/2} \frac{1}{2} \sqrt{\frac{4 + h_W}{h^W}} = +\infty \) 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 \( p \to 0 \) leads to \( K_1^B = 0 \).

Now we can substitute for \( J_k(p) \) in (15), and using the conditions \( J_W(p) = 0 \) and
For occupation $W$ and $p < \hat{p}$:

$$J_W(p) = \frac{r + \gamma + \delta_W + \eta_W - \lambda}{\zeta_W} p \left( 1 - p \right)^{1/2} \cdot \int_\pi^p \left[ \theta_W p + \kappa_W K_B^W \xi_W (1 - \tau)^{-1} + c_W \right] \left( \frac{\tau}{1 - \tau} \right) \tau^{-3/2} d\tau \label{eq:16}$$

where $l_W = \frac{1}{\lambda} + \frac{2(r + \gamma + \delta_W + \eta_W - \lambda)}{\zeta_W}$, $c_W = \frac{1 - q}{r + \gamma + \delta_W + \eta_W} \left( a_w - b_w \right) - a_W + b_W$, $\theta_W = \frac{1 - q}{r + \gamma + \delta_W + \eta_W} \left( a_W - b_W \right)$, $\kappa_W = \frac{1 - q}{r + \gamma + \delta_W + \eta_W} \left( a_W - b_W \right)^2 - 1 + \eta_W$ and $\xi_W = \frac{1}{2} \left( 1 + \sqrt{1 + 4 + h_B} \right)$.

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

$$J_B(p) = \frac{r + \gamma + \delta_B + \eta_B - \lambda}{\zeta_B} p \left( 1 - p \right)^{1/2} \cdot \int_\pi^p \left[ \theta_B p + \kappa_B K_B^W \xi_B (1 - \tau)^{-1} + c_B \right] \left( \frac{\tau}{1 - \tau} \right) \tau^{-3/2} d\tau \label{eq:17}$$

where $l_B = \frac{1}{\lambda} + \frac{2(r + \gamma + \delta_B + \eta_B - \lambda)}{\zeta_B}$ and $c_B = \frac{1 - q}{r + \gamma + \delta_B + \eta_B} \left( a_w - b_w \right) - a_B + b_B$, $\theta_B = \frac{1 - q}{r + \gamma + \delta_B + \eta_B} \left( a_B - b_B \right)$, $\kappa_B = \frac{1 - q}{r + \gamma + \delta_B + \eta_B} \left( a_B - b_B \right)^2 - 1 + \eta_B$ and $\xi_B = \frac{1}{2} \left( 1 + \sqrt{1 + 4 + h_B} \right)$.

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, $p$, $\hat{p}$ and $\hat{\pi}$, as well as the 2 yet undetermined coefficients $K_W$ and $K_B$. We need 5 conditions to do so. Optimality of searching behavior while unemployed, $V_W(\hat{p}) = V_B(\hat{\pi})$, provides one of these conditions. In the online appendix we show that the remaining 4 conditions are given by continuity of the total value of the match ($V_i(\cdot) + J_i(\cdot)$) at $\hat{\pi}$, as well as continuity
of its first derivative, for each occupation. Straightforward derivations imply that they can be rewritten as follows:

\[
\lim_{p \to \bar{p}^-} J_W (p) = \lim_{p \to \bar{p}^+} J_W (p) \tag{18}
\]

\[
\lim_{p \to \bar{p}^-} J_B (p) = \lim_{p \to \bar{p}^+} J_B (p) \tag{19}
\]

\[J_W (\bar{p}) = J_B (\bar{p}) \tag{20}\]

\[
\lim_{p \to \bar{p}^-} (V'_W (p) + J'_W (p)) = \lim_{p \to \bar{p}^+} (V'_W (p) + J'_W (p)) \tag{21}\]

\[
\lim_{p \to \bar{p}^-} (V'_B (p) + J'_B (p)) = \lim_{p \to \bar{p}^+} (V'_B (p) + J'_B (p)) \tag{22}\]

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 online appendix, the resulting solution is unique.

\section{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 \(p\) and \(z_i(p)\) denote 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 \(p \geq \underline{p}\) and \(p \neq \overline{p}\) is given by:

\[
0 = \frac{df_W (p)}{dt} = \frac{d^2}{dp^2} \left[ \frac{1}{2} \nu_W p^2 (1 - p)^2 f_W (p) \right] - \delta_W f_W (p) - \gamma f_W (p) + \lambda z_W (p) \tag{23}
\]

\[-\eta_W \lambda f_W (p) I \{p < \bar{p}\} + \eta_B \lambda f_B (p) I \{\overline{p} \leq p \leq \overline{p}\}\]

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, whereas 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 $p \neq \bar{p}$ is given by:

\[
0 = \frac{df_B(p)}{dt} = \frac{d^2}{dp^2} \left[ \frac{1}{2} \zeta_B^2 (1 - p)^2 f_B(p) \right] - \delta_B f_B(p) - \gamma f_B(p) + \lambda_B(p) \quad (24)
\]

\[-\eta_B \lambda_B(p) I \{ \bar{p} \leq p \} + \eta_W \lambda_W(p) I \{ p \leq p < \bar{p} \} \]

In order to solve for $f_W(\cdot)$ and $f_B(\cdot)$ we first need to solve for the population density of unemployed workers in each occupation $i \in \{W, B\}$, $z_i(\cdot)$. 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 $p \geq \hat{p}$ and $p \neq \bar{p}$:

\[
\delta_W f_W(p) + \delta_B f_B(p) + \gamma g(p) = \lambda z_W(p) + \gamma z_W(p) \quad (25)
\]

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, whereas 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 $p < \hat{p}$ and $p \neq \bar{p}$:

\[
\delta_W f_W(p) + \delta_B f_B(p) + \gamma g(p) = \lambda z_B(p) + \gamma z_B(p) \quad (26)
\]

We then use (25) and (26) to solve out for $z_W(p)$ and $z_B(p)$ respectively and after substituting them into the two forward equations (23) and (24), we derive the following system of differential equations:
\[
\frac{d^2}{dp^2} \left[ \frac{1}{2} \zeta_W^2 \left( 1 - p \right)^2 f_W(p) \right] - \gamma \frac{\delta_W + \lambda + \gamma}{\lambda + \gamma} f_W(p) + \frac{\lambda \delta_B}{\lambda + \gamma} f_B(p) \\
+ \frac{\gamma \lambda}{\lambda + \gamma} g(p) - \eta_W f_W(p) I \{ p < \bar{p} \} + \eta_B f_B(p) I \{ \bar{p} \leq p \leq \bar{p} \} = 0
\]

\[
\frac{d^2}{dp^2} \left[ \frac{1}{2} \zeta_B^2 \left( 1 - p \right)^2 f_B(p) \right] - \gamma \frac{\delta_B + \lambda + \gamma}{\lambda + \gamma} f_B(p) + \frac{\lambda \delta_W}{\lambda + \gamma} f_W(p) \\
+ \frac{\gamma \lambda}{\lambda + \gamma} g(p) - \eta_B f_B(p) I \{ \bar{p} \leq p \} + \eta_W f_W(p) I \{ p \leq p < \bar{p} \} = 0
\]

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

\[
\lambda \int_{\bar{p}}^{p} z_W(p) \, dp + \lambda \int_{\bar{p}}^{1} z_W(p) \, dp + \lambda z_W(p) + \eta_B \lambda \int_{\bar{p}}^{p} f_B(p) \, dp \\
= (\delta_W + \gamma) \int_{p}^{1} f_W(p) \, dp + \frac{1}{2} \zeta_W^2 \left( 1 - p \right)^2 f_W'(p) + \eta_W \lambda \int_{\bar{p}}^{p} f_W(p) \, dp
\]

\[
\int_{\bar{p}}^{1} f_W(x) \, dx \leq 1 < \infty
\]

\[
f_W(p) = 0
\]

\[
\lim_{\varepsilon \to 0} f_W(\bar{p} + \varepsilon) = f_W(\bar{p})
\]

\[
\lim_{\varepsilon \to 0} f_W(\bar{p} + \varepsilon) = f_W(\bar{p})
\]

\[
\lim_{\varepsilon \to 0} f_W'(\bar{p} + \varepsilon) = f_W'(\bar{p})
\]

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 \( \lambda \). At \( p = \bar{p} \), the endogenous exit of workers from occupation
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 \( p = \bar{p} \), 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 online 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 \( p \in [0, \hat{p}] \):

\[
\begin{align*}
 f_B (p) & = C_2^B p^{-1-q_B} (1 - p)^{q_B-2} - \frac{d_B}{\sqrt{c_B (4 + c_B)}} p^{q_B-2} (1 - p)^{1-q_B} \int_0^p \tau^{\psi_1-q_B} (1 - \tau)^{q_B+\psi_2-1} d\tau \\
 & + \frac{d_B}{\sqrt{c_B (4 + c_B)}} p^{-1-q_B} (1 - p)^{q_B-2} \int_0^p \tau^{q_B+\psi_1-1} (1 - \tau)^{\psi_2-q_B} d\tau
\end{align*}
\]  

(27)

for \( p \in (\bar{p}, \hat{p}) \):

\[
\begin{align*}
 f_B (p) & = C_3^B p^{q_B-2} (1 - p)^{1-q_B} + C_4^B p^{-1-q_B} (1 - p)^{q_B-2} \\
 & + \frac{1}{\sqrt{c_B (4 + c_B)}} p^{q_B-2} (1 - p)^{-1-q_B} [C_6^W m_B \int_p^{\hat{p}} \tau^{-\frac{1}{2} - \kappa_B-q_B} (1 - \tau)^{-\frac{1}{2} + \kappa_B+q_B} d\tau ] \\
 & - C_6^W n_B \int_p^{\hat{p}} \tau^{-\frac{1}{2} + \kappa_B-q_B} (1 - \tau)^{-\frac{3}{2} - \kappa_B+q_B} d\tau + d_B \int_p^{\hat{p}} \tau^{\psi_1-q_B} (1 - \tau)^{\psi_2+q_B-1} d\tau ] \\
 & - \frac{1}{\sqrt{c_B (4 + c_B)}} p^{-1-q_B} (1 - p)^{q_B-2} [C_6^W m_B \int_p^{\hat{p}} \tau^{-\frac{3}{2} - \kappa_B+q_B} (1 - \tau)^{-\frac{1}{2} + \kappa_B-q_B} d\tau ] \\
 & - C_6^W n_B \int_p^{\hat{p}} \tau^{-\frac{3}{2} + \kappa_B+q_B} (1 - \tau)^{-\frac{1}{2} - \kappa_B-q_B} d\tau + d_B \int_p^{\hat{p}} \tau^{\psi_1+q_B-1} (1 - \tau)^{\psi_2-q_B} d\tau ]
\end{align*}
\]  

(28)
and for $p \in [\hat{p}, \bar{p})$:

$$f_B (p) = C_B^5 \left( \frac{1}{p} (1 - p) \right)^{\frac{3}{2}} \left( \frac{1 - p}{p} \right)^{\kappa W} - \left( \frac{1 - p}{p} \right)^{2\kappa W} \left( \frac{p}{1 - p} \right)^{\kappa W}$$  \hspace{1cm} (29)$$

where $C_B^2, C_B^3, C_B^4, C_B^5$ and $C_B^6$ are undetermined coefficients and $c_B = \frac{c_B^2 (\lambda + \gamma)}{2 \gamma (\delta_B + \lambda + \gamma)}$, $q_B = \frac{1}{2} - \frac{1}{2} \sqrt{4 + c_B / c_B^2}$, $d_B = -\frac{\lambda}{(\delta_B + \lambda + \gamma) B(\psi_1, \psi_2)}$, $\kappa_W = \frac{1}{4} + \frac{2(\delta_B + \gamma + \eta_B \lambda)}{\zeta_B^2}$, $\kappa_B = \frac{1}{4} + \frac{2(\delta_B + \gamma + \eta_B \lambda)}{\zeta_W}$, $n_B = -\frac{\lambda (\delta_W + \gamma \lambda + \eta_W \gamma)}{\gamma (\delta_B + \lambda + \gamma)}$, and $m_B = -n_B \left( \frac{p}{1 - p} \right)^{2\kappa_B}$.

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

In the online appendix we also show that 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:

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

where:

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

As $p$ approaches zero, we drop all higher terms and we are left with:\footnote{Formally, since $\frac{1}{2} \left( 1 + \sqrt{\frac{4 + h_B}{\eta_B}} \right) > 1$, 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.}

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

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 $c_0 = q \lambda (1 - q) \frac{a_B^b - b_u}{r + \gamma + \delta_B + q \lambda} + q a_B^b + (1 - q) b_u$. 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 $\gamma > \frac{\lambda_2}{\lambda_1 \lambda + \lambda_2 + \lambda_1}$). The first case is the high-ability, low-ability workers case ($a_B^w - a_B^b = \frac{r+\gamma+\delta_B+q\lambda}{q(r+\gamma+\delta_B+q\lambda+\lambda(1-q))} > 0 \Rightarrow a_B^w > a_B^b$). 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 ($a_B^w - a_B^b = -\frac{r+\gamma+\delta_B+q\lambda}{q(r+\gamma+\delta_B+q\lambda+\lambda(1-q))} < 0 \Rightarrow a_B^w < a_B^b$), 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 $a_W^w - a_W^b = \frac{r+\gamma+\delta_W+q\lambda}{q(r+\gamma+\delta_W+q\lambda+\lambda(1-q))}$.

References


