Heterogeneity and Learning in Labor Markets[1]

Simon D. Woodcock[2]
Simon Fraser University
simon_woodcock@sfu.ca

May 2010

[1]This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01~AG018854, and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the author(s) and do not represent the views of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov http://lehd.dsd.census.gov).

[2]I especially thank John M. Abowd, David Easley, George Jakubson, Robert Shimer, Dale Mortensen, James Heckman, Martha Stinson, members of the LEHD program staff, the editor, and two anonymous referees for helpful comments on earlier versions of this paper. I also thank seminar participants at the CEA and CESG meetings, the University of Chicago, Cornell University, the North American winter meeting of the Econometric Society, the Federal Reserve Bank of Atlanta, McGill University, the University of Montreal, the NBER Summer Institute, Queen's Uni-

Abstract

I develop an equilibrium matching model where heterogeneous workers and firms learn about match quality and bargain over wages. The model generalizes Jovanovic (1979) to the case of heterogeneous workers and firms. Equilibrium wage dispersion arises due to productivity differences between workers, technological differences between firms, and heterogeneity in beliefs about match quality. Under a simple CRS technology, the equilibrium wage is additively separable in worker- and firm-specific components, and in the posterior mean of beliefs about match quality. This parallels the "person and firm effects" empirical specification of Abowd et. al. (1999) and others. The model predicts a negative correlation between estimated person and firm effects, which is consistent with most previous empirical evidence. I estimate the equilibrium wage function and test the model's empirical predictions using linked employer-employee data from the US Census Bureau. I find empirical support for many of the model's predictions, and estimate that dispersion in beliefs about match quality explains over 20 percent of observed earnings variation.

versity, RAND, Simon Fraser University, SOLE meetings, Texas A&M University, the University of Washington, the University of Western Ontario, and the University of Wisconsin-Madison.

1 Introduction

It is well known that observationally identical workers often earn very different wages and have heterogeneous employment histories. Likewise, otherwise similar firms frequently pay dissimilar wages and exhibit great heterogeneity in turnover. We have yet to fully understand this enormous variety of outcomes. A convincing explanation of why similar workers earn different wages, and how this is related to heterogeneity in job duration, unemployment, and the like, is central to our understanding of labor markets, and important for labor market policy.

Early work sought to explain wage differences across workers on the basis of variation in human capital and the non-pecuniary aspects of jobs. But observable characteristics of workers and firms usually only explain about 30 percent of wage variation. Attempts to explain the residual component of wage variation, often called *wage dispersion*, have proceeded along several dimensions. Search and matching models reveal that labor market frictions are one cause of equilibrium wage dispersion and unemployment. Learning models provide an explanation for wage dynamics and separation behavior, and show that heterogeneous beliefs about match productivity are another source of wage dispersion. And recent empirical work using linked employer-employee data shows that wage dispersion can be decomposed into a component attributable to unobserved characteristics of the worker – a "person effect" – and a component attributable to unobserved characteristics of the employer – a "firm effect." Each of these provides a partial explanation for the diversity of labor market outcomes. To date, however, they have remained distinct. This paper demonstrates that they are complementary. Together, they provide a comprehensive explanation for heterogeneity in labor market outcomes, and important new insights.

The paper has both theoretical and empirical components. In the first half of the paper, I develop an equilibrium matching model where heterogeneous workers and firms learn about match quality and bargain over wages. The main theoretical innovation is to embed learning about match quality in a Mortensen-Pissarides style equilibrium matching model with heterogeneous agents. This delivers novel insights into the relationship between worker and firm heterogeneity, wages, and separation behavior.

The matching model generalizes the canonical Jovanovic (1979) model to the case of heterogeneous workers and firms. That model provides an explanation for job duration and turnover: matches last as long as agents believe the match is highly productive. If they learn that match productivity is low, they prefer to separate. However, models with homogeneous agents can not explain why some workers experience consistently longer job duration than others, and why some firms experience less turnover than others. In the model presented here, workers and firms vary in their marginal productivity. The productivity of a worker-firm match depends on worker and firm productivity, as well as the quality of the match between them. Workers and firms learn the value of match quality slowly by observing production outcomes. Like the Jovanovic model, they terminate the employment relationship if beliefs about match quality fall below a threshold. A key result is that the reservation value is decreasing in both the worker's and the firm's productivity. Consequently, more productive workers experience longer average job duration than less productive workers. Likewise, firms with more productive technologies experience less turnover than less productive firms. This is consistent with empirical evidence, and provides an explanation for heterogeneity in job duration and turnover.

Distinguishing between worker, firm, and match heterogeneity is an important departure from earlier research. It recognizes that workers are differently able, and hence some are more productive on average than others. Likewise, it recognizes that firms operate different production technologies, and consequently employee productivity varies across firms. It also recognizes that not all workers are equally suited to all production technologies. As a consequence, two workers that are equally able may be differently productive in a given firm, simply because one is well matched to the firm's production technology and the other is not. It is easy to construct real world examples of this phenomenon. For instance, two equally able academics may have different proclivities for teaching and research. One will thrive in a university that emphasizes research while the other's productivity suffers. The reverse will be true in a university that emphasizes teaching.

The matching model yields an equilibrium wage function that is additively-separable in a worker-specific component, a firm-specific component, and the mean of beliefs about match quality. The worker- and firm-specific components reflect worker and firm productivity, adjusted for bargaining strength and discounting. This result is important for several reasons. First, it provides a rich explanation for wage dispersion: equilibrium dispersion arises due to productivity differences between workers, technological differences between firms, and due to heterogeneity in beliefs about match quality. Second, the additively-separable structure parallels the empirical "person and firm effects" specification of Abowd et al. (1999, AKM, hereafter) and others. This specification typically explains about 90 percent of observed wage variation, but provides no formal economic interpretation of what the person and firm effects actually measure. Thus the matching model contributes to the empirical literature by providing a theoretical context for the AKM specification.

The model also provides insight into the sorting behavior of workers and firms. The optimal separation policy implies that worker-firm matches are negatively assortative. Consequently, the matching model predicts a negative correlation between estimated person and firm effects. This is in fact what most prior empirical studies have found, but it has been considered something of an empirical puzzle. The model provides an intuitive explanation for this finding. Because high-productivity workers have a high opportunity cost of unemployment, they are willing to match with low-productivity firms. Likewise, when highly-productive firms have an unfilled vacancy, they forego more output than low-productivity firms do. They are consequently willing to match with less productive workers to avoid leaving a vacancy unfilled. The result is equilibrium mismatch.

To keep the model tractable, I make some important simplifying assumptions. First and foremost, I assume that workers only search for jobs when unemployed. Although others have shown that on-the-job search is an important source of wage dispersion and employment mobility,[1] our main objective is to investigate how worker and firm heterogeneity interact with learning to determine wages and mobility. Thus I focus on a single source of endogenous mobility (learning) for clarity and simplicity. Second, I assume that agents observe each other's type upon meeting in the matching market. Others have investigated how learning about ability contributes to wage growth,[2] but such models generally have little to say about turnover. We focus on learning about match quality, rather than learning about ability, because the Jovanovic (1979) model has long been a benchmark model for worker turnover; and because it allows us to explore how learning jointly determines turnover and wage dynamics, rather than wage dynamics alone. Generalizing the model to the case where agents learn about ability *and* match quality substantially complicates the analysis, so we leave such considerations for future research.

In the second half of the paper, I estimate structural parameters of the matching model using linked employer-employee data, and test a variety of its predictions. The main econometric innovation is to estimate a random effect specification of the equilibrium wage function that allows a completely unrestricted within-match error covariance. This generalizes the fixed effect estimator that AKM and others have used. I then fit the specific error covariance structure implied by learning about match quality to the unrestricted

[1] See Mortensen (1994), Postel-Vinay and Robin (2002), Cahuc et al. (2006) and others.

[2] e.g., Farber and Gibbons (1996), Altonji and Pierret (2001), Gibbons et al. (2005), and Lange (2007).

estimate, and test whether it is consistent with the data. Though the data reject the learning structure, they do not do so resoundingly.

The empirical results shed light on the relative importance of different sources of wage dispersion. I find that personal heterogeneity explains over 50 percent of earnings variation, and dispersion in beliefs about match quality explains more than 20 percent. Employer heterogeneity explains a more modest 6.3 percent.

The remainder of the paper is structured as follows. I begin by briefly reviewing the related literature. I then present the matching model in Section 2, and develop the empirical specification in Section 3. I describe the data in Section 4, present the empirical results in Section 5, and conclude in Section 6.

1.1 Related Literature

The model presented here brings together three distinct literatures: that on search and matching with heterogeneous agents, that on learning in labor markets, and an empirical literature that seeks to explain wage dispersion using linked employer-employee data.

1.1.1 Search and Matching with Heterogeneous Agents

In general, the search and matching literature has focused on economies with heterogeneous workers and jobs.[3] In the typical model, firms employ only a single worker. There is therefore no distinction between heterogeneity at the level of the firm and at the level of the worker-firm match. In contrast, I model an economy where firms employ many workers, and distinguish between productive heterogeneity at the firm, which affects all employees, and productive heterogeneity that is specific to a worker-firm match. Postel-Vinay and Robin (2002) also consider an environment where firms employ many workers, but their workers and firms do not learn about match quality.

Recently, interest has focused on conditions under which "good" workers sort into "good" firms. Shimer and Smith (2000) develop conditions under which assortative matching arises in the presence of search frictions. Shimer (2005) considers the optimal assignment of workers to jobs in the presence of coordination frictions. In either case, positively assortative matching arises if

[3] Examples include Stern (1990), Sattinger (1995), Shimer and Smith (2000), and Shimer and Smith (2001). Albrecht and Vroman (2002), Gautier (2002), and Kohns (2000) develop models with exogenous heterogeneity on one side of the market, and endogenous heterogeneity on the other.

production exhibits sufficient complementarity between worker and firm types. Here, we consider a constant returns production technology that does not allow for complementarities between the quality of workers, firms, and matches. This implies that good workers and good firms have a comparative advantage in producing rather than searching for new matches. Good firms (workers) will therefore accept comparatively bad matches and will accept to be matched with relatively bad workers (firms). That is, worker-firm matches are negatively assortative in equilibrium. This implies a negative correlation between empirical person and firm effects, which is consistent with many previous empirical studies.[4]

1.1.2 Learning in Labor Markets

The learning literature has focused primarily on wage and turnover dynamics. The seminal Jovanovic (1979) model considered the case where identical workers and firms learn about the quality of a match. Flinn (1986) and Moscarini (2003) generalize the canonical model to the case of heterogeneous workers. Nagypal (2007) considers the case where agents learn about match quality and workers learn by doing. Harris and Holmstrom (1982) and Farber and Gibbons (1996) present models where workers and firms learn about a worker's unobservable ability, which is correlated with observable characteristics. Gibbons et al. (2005) extend this framework to the case of an economy with heterogeneous sectors, and where workers exhibit comparative advantage in some sectors. Felli and Harris (1996) present a model where workers learn about their aptitude for firm-specific tasks. Moscarini (2005) embeds learning in a Mortensen-Pissarides style matching model, but where agents are homogeneous. None of these earlier works consider the case of worker, firm, and match heterogeneity.

1.1.3 Estimating Person and Firm Effects

I estimate both fixed and random effect specifications of the equilibrium wage function. All prior studies are based on a fixed effect estimator of person and firm effects. These include AKM, Abowd et al. (2002, ACK hereafter), Abowd et al. (2003), Barth and Dale-Olsen (2003), Abowd et al. (2004), Gruetter

[4] Abowd et al. (2002) and Abowd et al. (2004) report a negative correlation in French data, and approximately zero correlation in American data. Gruetter and Lalive (2004) find a negative correlation in Austrian data, Barth and Dale-Olsen (2003) find a negative correlation in Norwegian data, and Andrews et al. (2004) find no significant correlation in German data.

and Lalive (2004), Abowd et al. (2005), Andersson et al. (2005), and Andrews et al. (2008). The empirical specification considered by these authors, which I refer to generically as the AKM model, is

$$y_{ijt} = \mu + x'_{it}\beta + \theta_i + \psi_j + \varepsilon_{ijt} \tag{1}$$

where i indexes workers, j indexes firms, t indexes calendar time, y_{ijt} is a measure of earnings, μ is the grand mean, x_{it} is a vector of covariates, β is a parameter vector, θ_i is the person effect, ψ_j is the firm effect, and ε_{ijt} is statistical error.

Economists usually prefer fixed effect estimators over alternatives that treat unobserved heterogeneity as random. Statisticians, on the other hand, generally prefer random effect specifications, since they yield estimates of the unobserved heterogeneity (θ_i, ψ_j) that have better sampling properties; see Robinson (1991). Furthermore, it is well known (at least among statisticians) that the fixed effect estimator is a limiting case of the random effect estimator, see e.g., Searle et al. (1992) or McCulloch and Searle (2001). Of course economists usually prefer the fixed effect estimator because β is the object of interest in most applications. Identifying β via a random effect estimator requires that observables (x_{it}) are uncorrelated with unobservables. Here, however, our main interest is to estimate parameters of the distribution of θ_i and ψ_j, not β.

2 The Matching Model

The economy is populated by a continuum of infinitely-lived workers of measure one. There is a continuum of firms of measure ϕ. All agents are risk neutral and share the common discount factor $0 < \rho < 1$. Time is discrete.

In each period, workers are endowed with a single indivisible unit of labor that they supply to production at home or at a firm. Workers vary in their marginal productivity when employed, denoted $a \in [a_0, a_1]$. Conceptually, a represents the worker's ability, motivation, and the like. I refer to a as worker quality or ability. Let

$$a \sim F_a \text{ iid across workers} \tag{2}$$

where F_a is a probability distribution with zero mean, known to all agents. Worker quality is exogenous, known to the worker, and observed by the firm when the worker and firm meet. Unemployed workers receive income $h \in \mathbb{R}$

from home production.[5] For simplicity, h includes all search costs, the value of leisure, and the like. Workers maximize the expected present value of wages.

Firms employ many workers. They operate in a competitive output market and produce a homogeneous good whose price is normalized to one. Firms can only produce output when matched with workers. They seek to maximize the expected net revenues of a match: the expected value of output minus a wage payment to the worker.

Firms vary in their technology, which determines the marginal productivity of all their employees, denoted $b \in [b_0, b_1]$. Let

$$b \sim F_b \text{ iid across firms} \tag{3}$$

where F_b is a probability distribution known to all agents, and with zero mean. I call b firm quality. Firms know their own value of b, and it is observed by the worker when the worker and firm meet. Note b is exogenous. Firms open vacancies at per unit cost k_0, and they incur cost $\kappa(l)$ to hire l workers.[6] Assume κ is continuous, increasing, and convex.

The marginal productivity of a type a worker when employed at a type b firm also depends on a worker- and firm-specific interaction that I call match quality, denoted c. "Good" matches are more productive than "bad" ones, all else equal. Let

$$c \sim N\left(0, \sigma_c^2\right) \text{ iid across matches.} \tag{4}$$

The normality assumption follows Jovanovic (1979) and others. It yields a convenient closed form for beliefs about match quality.

Match quality c is a pure experience good. It is unobserved by either the worker or the firm. They learn its value slowly. When the worker and firm meet, they observe the noisy signal $x = c + z$ where

$$z \sim N\left(0, \sigma_z^2\right) \text{ iid across matches.} \tag{5}$$

The worker and firm's initial beliefs about c are based on a common prior and the signal x. They subsequently update their beliefs about c on the basis of output realizations. Prior beliefs and the updating process are discussed in Section 2.1. Note that information is incomplete, since c is unobserved, but is symmetric. That is, the worker and firm both know a and b, and observe the

[5] We can let h vary across individuals without changing any key theoretical results. However it changes the interpretation of the person-specific component of wages (Section 2.2.3), the person-specific term in the reservation level of beliefs (Section 2.2.4), and complicates the comparative statics in Section 2.3.

[6] Firm-specific vacancy-opening costs do not affect any of the main results, but introduce an additional source of firm-level heterogeneity in hiring and employment growth.

same signals about c. They therefore share common beliefs about c at every point in time.

Output is produced according to the constant returns to scale production function:

$$q_\tau = \mu + a + b + c + e_\tau \tag{6}$$

where τ indexes tenure (the duration of the match), μ is the grand mean of productivity (known to all agents), and e_τ is a match- and tenure-specific idiosyncratic shock. Let

$$e_\tau \sim N\left(0, \sigma_e^2\right) \text{ iid across matches and tenure.} \tag{7}$$

Note the Jovanovic (1979) production technology is a (continuous time) special case of (6) where workers and firms are homogeneous. Note also that there are no aggregate shocks to productivity, and no human capital accumulation over the life cycle.[7] Since a, b, and μ are known, agents extract the noisy signal of match quality $c + e_\tau$ from production outcomes q_τ.

Following Flinn (1986), I assume that q_τ is bounded. This implies that the random variables c, z, and e_τ have bounded support. Thus the distributional assumptions (4), (5), and (7) are approximate.

Unemployed workers are matched to firms with open vacancies. Search is undirected. The total number of matches formed each period is $m(u, v)$, where u is the number of unemployed workers and v is the number of open vacancies. Both u and v are determined endogenously. Assume m is non-decreasing in u and v. The probability that a randomly selected unemployed worker will be matched to a firm this period is $\pi \equiv m(u, v)/u$. Similarly, the probability that a randomly selected vacancy will be filled is $\lambda \equiv m(u, v)/v$. With a large number of workers and firms, all agents take u and v as given.

As discussed below, a match between a worker and firm terminates endogenously when their point estimate of match quality falls below a threshold value. In addition, matches terminate with exogenous probability $\delta > 0$ in each period.

[7] Introducing a publicly-observable aggregate shock to productivity is straightforward. The same is true of a deterministic trend in individual productivity (i.e., an "experience effect") provided it is observable by the worker and firm. I abstract from these considerations since they complicate the exposition considerably – both require additional notation and an index of calendar time – but without loss of generality. Indeed, the production function (6) can be considered net of additive aggregate shocks and deterministically accumulated human capital. The same is true of the equlibrium wage w_τ in (26) and the net value of output $q_\tau - w_\tau$. That is, in the more general model, the equilibrium wage (see Proposition 1) remains additively separable in person- and firm- specific components and in the posterior mean of beliefs, and is linear and additively separable in the productivity shock and an experience effect.

I restrict attention to the steady state. The economy is in steady state when the end-of-period distribution of type a workers employed at type b firms and unemployed is constant. The various flow-balance equations that characterize the steady state are given in Appendix B. These imply that the steady state values of u and v are constant. Hence so are the steady state values of λ and π.

Within-period timing is as follows:

1. With probability π, unemployed workers are randomly matched to a firm with an open vacancy. Upon meeting, agents observe a, b, and the signal x.

2. Matched workers and firms decide whether to continue the match. The decision is based on all current information about the match: a, b, and current beliefs about c. The current period wage w_τ is simultaneously determined by a Nash bargain.

3a. If agents decide to terminate the match, the worker enters unemployment and receives h. There are no firing costs.

3b. If agents decide to continue the match, the negotiated wage is paid to the worker and output q_τ is produced. (Assume that reputational considerations preclude agents from reneging on the agreed-upon wage payment.) Agents update their beliefs about c.

4. Exogenous separations occur with probability δ.

5. Firms open new vacancies v.

2.1 Beliefs About Match Quality

Assume agents' prior beliefs about a, b, c, z, and e_τ are rational. That is, they are governed by equations (2), (3), (4), (5), and (7). Agents update their beliefs about match quality using Bayes' rule when they acquire new information, i.e., upon observing the signal x and production outcomes q_τ.

After observing x, worker and firm posterior beliefs about c are normally distributed with mean m_1 and variance s_1^2 where

$$m_1 = x\left(\frac{\sigma_c^2}{\sigma_c^2+\sigma_z^2}\right) \quad \text{and } s_1^2 = \frac{\sigma_c^2\sigma_z^2}{\sigma_c^2+\sigma_z^2}. \tag{8}$$

In each subsequent period that the match continues, the worker and firm extract the signal $c + e_\tau$ from observed output q_τ. Hence at the beginning of the τ^{th} period of the match (that is, after observing $\tau-1$ production outcomes), their posterior beliefs about match quality are normally distributed with mean m_τ and variance s_τ^2, where

$$m_\tau = \left(\frac{m_{\tau-1}}{s_{\tau-1}^2} + \frac{c+e_{\tau-1}}{\sigma_e^2}\right) / \left(\frac{1}{s_{\tau-1}^2} + \frac{1}{\sigma_e^2}\right) \quad \text{and} \quad \frac{1}{s_\tau^2} = \frac{1}{s_{\tau-1}^2} + \frac{1}{\sigma_e^2}. \tag{9}$$

Equation (9) demonstrates that s_τ^2 evolves deterministically and does not depend on the value of the signals received, and that $s_\tau^2 < s_{\tau-1}^2$ for each $\tau > 0$. We also see that m_τ is a precision-weighted average of the prior mean $m_{\tau-1}$ and the most recent signal $c + e_{\tau-1}$. Since the precision of signals $(1/\sigma_e^2)$ is constant but the precision of beliefs $(1/s_\tau^2)$ increases with tenure, each new signal is given successively smaller weight in the update. Asymptotically, beliefs converge to unit mass at true match quality. That is, $\lim_{\tau\to\infty} m_\tau = c$ and $\lim_{\tau\to\infty} s_\tau^2 = 0$. This is a standard result for Bayesian learning with rational priors.

In what follows, it will be of interest to describe the cross-sectional distribution of beliefs. It is easy to show that $m_\tau \sim N(0, V_\tau)$ in the population, where

$$V_\tau = s_\tau^2 \sigma_c^2 \left(\frac{1}{\sigma_z^2} + \frac{\tau-1}{\sigma_e^2}\right). \tag{10}$$

With a little algebra, one can also show $V_{\tau+1} > V_\tau$ for all $\tau > 0$. Another standard result for Bayesian learning with normal priors and signals is

$$m_p | m_\tau \sim N(m_\tau, v_p) \tag{11}$$

$$v_p = \frac{s_\tau^4 (p-\tau)}{s_\tau^2 (p-\tau) + \sigma_e^2} \tag{12}$$

for any $p > \tau$. Equation (11) implies m_τ is a martingale. Conditional on current information, the expected value of any future realization of m_τ equals its current value.

2.2 Match Formation, Duration, and Wages

In each period, wages are determined by a Nash bargain between the worker and the firm. They divide the expected match surplus. They take expectations with respect to current beliefs about match quality, given a and b. The equilibrium wage therefore maps current information about the match (a, b, m_τ, s_τ^2) into a payment from the firm to the worker. Because the Nash bargain is efficient, the match only continues if the expected surplus is non-negative. Otherwise, the worker and firm both prefer to separate.

Let J_τ denote the worker's value of employment at tenure τ. Let U denote the value of the worker's outside option (unemployment). Let Π_τ denote the firm's value of employment at tenure τ, and let V denote the value of the firm's outside option (a vacancy). In the steady state, U and V are constant. At tenure τ, the match continues if and only if

$$J_\tau + \Pi_\tau \geq U + V. \tag{13}$$

When (13) is satisfied, the equilibrium wage w_τ solves the Nash bargaining condition

$$(1-\gamma)(J_\tau - U) = \gamma(\Pi_\tau - V) \tag{14}$$

where γ is the worker's exogenous share of match surplus.

2.2.1 The Worker's Value of Employment and Unemployment

The worker's expected value of employment at wage w_τ is

$$J_\tau = w_\tau + \rho(1-\delta) E_\tau [\max\{J_{\tau+1}, U\}] + \rho\delta U \tag{15}$$

where E_τ denotes the expectation taken with respect to tenure τ information, (a, b, m_τ, s_τ^2). For what follows, it is convenient to rewrite J_τ net of the value of unemployment:

$$J_\tau - U = w_\tau + \rho(1-\delta) E_\tau [\max\{J_{\tau+1} - U, 0\}] - (1-\rho) U. \tag{16}$$

The steady state value of being unemployed today and behaving optimally thereafter is

$$U = h + \rho\pi \int_{b_0}^{b_1} J_0 dF_b^* + \rho(1-\pi) U \tag{17}$$

where F_b^* is the steady state distribution of firm types among open vacancies (see Appendix B), and

$$J_0 = E_0 [\max\{J_1, U\}] \tag{18}$$

is the expected value of employment before the initial signal of match quality is observed.

2.2.2 Vacancies and The Firm's Value of Employment

The firm's value of employing a worker at wage w_τ is today's expected net revenues plus the discounted expected value of employment next period. Thus,

$$\begin{aligned}\Pi_\tau &= E_\tau [q_\tau] - w_\tau + \rho(1-\delta) E_\tau [\max\{\Pi_{\tau+1}, V\}] + \rho\delta V \\ &= \mu + a + b + m_\tau - w_\tau + \rho(1-\delta) E_\tau [\max\{\Pi_{\tau+1}, V\}] + \rho\delta V \end{aligned} \tag{19}$$

so that

$$\Pi_\tau - V = \mu + a + b + m_\tau - w_\tau + \rho\left(1-\delta\right)E_\tau\left[\max\left\{\Pi_{\tau+1} - V, 0\right\}\right] - \left(1-\rho\right)V. \quad (20)$$

The production technology (6) implies that the firm's employees produce independently of one another. As a consequence, the firm's decision to open vacancies is static. The number of hires today has no dynamic consequences for future hiring or productivity. When a firm opens υ vacancies, the number l that are filled is a binomial process. The number of vacancies opened by a type b firm in each period therefore solves

$$\max_{\upsilon \in \mathbb{N}} \sum_{l=0}^{\upsilon} \binom{\upsilon}{l} \lambda^l \left(1-\lambda\right)^{\upsilon - l} \left[l \int_{a_0}^{a_1} \Pi_0 dF_a^* - \kappa\left(l\right)\right] - k_0 \upsilon \quad (21)$$

where F_a^* is the steady state distribution of unemployed worker types defined in Appendix B, and

$$\Pi_0 = E_0\left[\max\left\{\Pi_1, V\right\}\right] \quad (22)$$

is the expected net revenues from a match before the signal x is observed.

Note that firm size (employment) is indeterminate. I derive the average steady state employment of a type b firm in Appendix B.[8] Increasing and convex hiring costs κ, however, coupled with per-unit vacancy-opening costs k_0, guarantee that the solution to (21) is well defined and the firm opens a finite number of vacancies in each period. In particular, since firms are free to open vacancies, they do so until V is bid down to zero. Equivalently, since hiring costs are sunk, terminating an employment relationship frees up no resources. Thus the alternative value of a vacancy is zero.

2.2.3 The Equilibrium Wage

With the value functions in hand, it is a simple matter to solve for the equilibrium wage. It takes a simple form, summarized in Proposition 1.

Proposition 1 (Equilibrium Wage) *At each tenure $\tau > 0$, the equilibrium wage w_τ is linear and additively separable in a person-specific component, a firm-specific component, and the posterior mean of beliefs about match quality; and is independent of s_τ^2.*

[8] As in Shimer and Smith (2000), Postel-Vinay and Robin (2002), Shimer (2005), and Cahuc et al. (2006), labor market frictions allow high- and low-productivity firms to coexist in equilibrium. Here, firms survive (i.e., have positive employment) as long as (13) is satisfied in at least one match. Given our assumption that b and c are independent, even very low productivity firms will be a good match for some workers, so that they have positive employment with positive probability.

Proof. Substituting (16) and (20) into the Nash bargaining condition (14) we obtain

$$\begin{aligned} & (1-\gamma)\left(w_\tau + \rho\left(1-\delta\right) E_\tau\left[\max\left\{J_{\tau+1} - U, 0\right\}\right] - \left(1-\rho\right) U\right) \\ = \; & \gamma\left(\mu + a + b + m_\tau - w_\tau + \rho\left(1-\delta\right) E_\tau\left[\max\left\{\Pi_{\tau+1} - V, 0\right\}\right]\right) \\ & -\gamma\left(1-\rho\right) V. \end{aligned} \tag{23}$$

Condition (14) implies

$$(1-\gamma) E_\tau\left[\max\left\{J_{\tau+1} - U, 0\right\}\right] = \gamma E_\tau\left[\max\left\{\Pi_{\tau+1} - V, 0\right\}\right] \tag{24}$$

and thus

$$(1-\gamma)\left(w_\tau - (1-\rho) U\right) = \gamma\left(\mu + a + b + m_\tau - w_\tau - (1-\rho) V\right). \tag{25}$$

Rearranging yields

$$w_\tau = \gamma\mu + \theta + \psi + \gamma m_\tau \tag{26}$$

where the worker specific component is

$$\theta = \gamma a + (1-\gamma)(1-\rho) U \tag{27}$$

and the firm-specific component is

$$\psi = \gamma b - \gamma (1-\rho) V = \gamma b \tag{28}$$

given the equilibrium condition $V = 0$. ■

The equilibrium wage function (26) has the same additively separable structure as the AKM empirical specification (1). In keeping with the empirical literature, I therefore refer to θ and ψ as person and firm effects, respectively. The firm effect measures the worker's share of the firm's contribution to match surplus. Rewriting equation (27) as $\theta = \gamma\left(a - (1-\rho) U\right) + (1-\rho) U$ demonstrates that the person effect is likewise the worker's share of his contribution to match surplus, plus compensation for forgoing his next-best alternative.

The Jovanovic (1979) equilibrium wage is a special case of (26). In that model, workers and firms are identical but matches are heterogeneous, and production occurs according to the continuous time analog of (6) with $a = b = 0$ for every worker and firm. The Jovanovic wage equals the expected marginal product of the match, which in that model is the posterior mean of beliefs about match quality. This result relies on the assumption that firms earn zero expected profit. Similar to Jovanovic's model, the equilibrium wage (26) is linear and additively separable in expected marginal product, $\mu + a + b + m_\tau$, and in the posterior mean of beliefs about match quality. To see that the Jovanovic (1979) equilibrium wage is a special case, note that when workers capture the entire match surplus, i.e., as $\gamma \rightarrow 1$ (so that firms earn zero expected profit), the equilibrium wage is $\lim_{\gamma\rightarrow 1} w_\tau = \mu + a + b + m_\tau$. That is, it converges to the expected marginal product of the match.

The wage function (26) implies rich equilibrium wage dispersion. The person effect, the firm effect, and learning about match quality all contribute. Identical workers earn different wages because of employment at heterogeneous firms and because of heterogeneity in beliefs about match quality. Identical firms pay different wages because they employ heterogeneous workers and because of heterogeneity in beliefs. Even identical workers employed in identical firms earn different wages because of dispersion in beliefs about match quality.

2.2.4 The Separation Decision

The separation decision is made jointly by the worker and firm. The match continues as long as the surplus is non-negative. To characterize the separation decision, it is useful to introduce the Bellman equation for the joint value of employment, W :

$$\begin{aligned} W\left(m_\tau, s_\tau^2\right) &= \max\left\{J_\tau + \Pi_\tau, U + V\right\} \\ &= \max\left\{\begin{array}{c} \mu + a + b + m_\tau \\ +\rho\left(1-\delta\right)E_\tau\left[W\left(m_{\tau+1}, s_{\tau+1}^2\right)\right] + \rho\delta U, U \end{array}\right\} \end{aligned} \quad (29)$$

given the equilibrium condition $V = 0$. I suppress W's dependence on a and b for notational simplicity, and because these quantities do not vary with tenure.

The following Proposition establishes uniqueness of the value function and its most important properties. The proof is in Appendix A.

Proposition 2 (Uniqueness) *There is a unique value function W that satisfies the Bellman equation (29). Furthermore, W is continuous, non-decreasing, and convex in m_τ.*

Workers and firms prefer to continue the employment relationship as long as (13) is satisfied. There are a number of equivalent ways of characterizing this decision in terms of state variables. The most convenient is in terms of beliefs about match quality, since the other state variables do not vary during the match. The main result is summarized in Proposition 3; see Appendix A for the proof.

Proposition 3 (Optimal Separation Policy) *At each tenure $\tau > 0$ and for given values of a and b, the optimal separation policy is characterized by a reservation value of beliefs about match quality, $\bar{m}_\tau$. Specifically, the optimal policy is to separate if $m_\tau < \bar{m}_\tau$, and continue if $m_\tau \geq \bar{m}_\tau$.*

The reservation level of beliefs about match quality is the value of m_τ at which workers and firms are indifferent between continuing the employment relation and terminating it. Thus $\bar{m}_\tau$ satisfies (13) with equality, or equivalently, equates the arguments of the max function in the Bellman equation. Thus $\bar{m}_\tau$ is implicitly defined by

$$\bar{m}_\tau = (1 - \rho\delta) U - \mu - a - b - \rho (1 - \delta) \bar{E}_\tau \left[W \left(m_{\tau+1}, s^2_{\tau+1} \right) \right] \tag{30}$$

where $\bar{E}_\tau$ denotes the expectation taken with respect to $\left(a, b, \bar{m}_\tau, s^2_\tau \right)$.

Proposition 4 characterizes how separation behavior evolves with tenure. Its proof is in Appendix A.

Proposition 4 (Monotonicity) *The reservation value of beliefs about match quality is monotone in tenure, i.e., $\bar{m}_{\tau+1} \geq \bar{m}_\tau$ for all $\tau > 0$.*

As in most Bayesian learning models, this result reflects the option value of employment. Early in the match, when beliefs about match quality are imprecise, workers and firms are willing to accept matches of low believed quality because their point estimate m_τ is imprecise. As the worker and firm acquire more information, their beliefs become increasingly precise. As a consequence, the worker and firm become increasingly selective about admissible values of match quality, and the reservation value increases. Asymptotically, $\lim_{\tau \to \infty} \bar{m}_\tau = [1 - \rho (1 - \delta)] U - \mu - a - b$.

2.3 Comparative Statics

There are several equivalent ways to characterize how separation behavior varies with worker and firm quality. Proposition 5 characterizes how $\bar{m}_\tau$ varies with worker and firm quality. The proof is in Appendix A. An appendicized lemma (Lemma 6) further establishes that $\partial\theta/\partial a > 0$ and $\partial\psi/\partial b > 0$, so Proposition 5 can be restated in terms of the empirical person and firm effects, instead of worker and firm quality, with no change in result.

Proposition 5 *At each tenure $\tau > 0$, the reservation value of beliefs about match quality is decreasing in worker and firm quality. That is,*

$$\frac{\partial \bar{m}_\tau}{\partial a} < 0, \quad \frac{\partial \bar{m}_\tau}{\partial b} < 0. \tag{31}$$

This result is fairly intuitive. Consider a small change in the firm's quality b. An increase in b raises the value of remaining in the match $(J_\tau + \Pi_\tau)$ without changing the value of terminating it $(U + V)$, and consequently makes the worker and firm less selective about match quality. That is, having found a "good" employer, the worker is less picky about whether or not it is a good match; and since all workers are highly productive at "good" firms, the firm is less picky about whether or not they are good matches.

Similar intuition explains why an increase in a reduces $\bar{m}_\tau$, with one complication: increasing a raises the worker's productivity not only in the current match, but in all matches. That is, the value of U is increasing in a (see Lemma 6 in Appendix A). Nevertheless, matching frictions ensure that increasing a raises the value of continuing the match more than the value of terminating it.[9] From the firm's perspective, having found a "good" employee, the firm is less picky about whether or not she is a good match.

Proposition 5 has obvious implications for the relationship between job duration and worker/firm quality. Expected job duration is decreasing in $\bar{m}_\tau$, and thus is increasing in worker and firm quality. By extension, it is also increasing in θ and ψ. This is consistent with stylized facts and we find corroborating empirical evidence in Section 5.

We have characterized the separation decision in terms of $\bar{m}_\tau$, which is the most natural way to analyze separation behavior *within* a match. Alternately, we can characterize the separation decision in terms of a reservation level of worker quality, $\bar{a}$, for given firm quality and beliefs about match quality; or symmetrically, in terms of a reservation level of firm quality, $\bar{b}$, for given values of a, m_τ, and s_τ^2. This is a more natural way to characterize how workers sort across firms. Like $\bar{m}_\tau$, $\bar{a}$ and $\bar{b}$ are defined by equating the two arguments of the max operator in the value function (29). That is, they are implicitly defined by:

$$\bar{a} = (1-\rho\delta)\,U - \mu - b - m_\tau - \rho\,(1-\delta)\,E_\tau\left[W\left(m_{\tau+1}, s_{\tau+1}^2\right)\right] \quad (32)$$
$$\bar{b} = (1-\rho\delta)\,U - \mu - a - m_\tau - \rho\,(1-\delta)\,E_\tau\left[W\left(m_{\tau+1}, s_{\tau+1}^2\right)\right] \quad (33)$$

where the expectation in (32) is taken with respect to $(\bar{a}, b, m_\tau, s_\tau^2)$, and the expectation in (33) is taken with respect to $\left(a, \bar{b}, m_\tau, s_\tau^2\right)$. Differentiating (32) and (33) reveals:[10]

$$\frac{\partial \bar{a}}{\partial b} < 0,\ \frac{\partial \bar{b}}{\partial a} < 0. \quad (34)$$

[9] That is, using the result of Lemma 6 in Appendix A, it is easy to show that $\partial J_\tau/\partial a > \partial U/\partial a$.

[10] The algebra is omitted but available on request. The method of proof parallels that of Proposition 5.

That is, holding beliefs about match quality constant, matches are negatively assortative: more productive firms are willing to match with less productive workers, and vice versa. This result is consistent with other models where production is additively separable in agents' types, e.g., search examples in Shimer and Smith (2000) and an assignment example in Shimer (2005). Negative assortative matching is also consistent with recent empirical evidence that finds a negative correlation between estimated person and firm effects.[11] The intuition is straightforward. There are no complementarities in production between worker, firm, and match quality, so good workers (good firms) have a comparative advantage in producing rather than searching for good matches. They will therefore accept to be matched with relatively bad firms (bad workers). Conversely, low-ability workers (low-productivity firms) have a very low opportunity cost of searching (a vacancy) and are consequently most willing to wait for a match with a good firm (good worker).

2.4 Discussion

Before turning to empirics, it is useful to discuss some of the matching model's other predictions about wage dynamics and job duration. We will look for the empirical counterparts to these in Section 5.

Since the person and firm effects do not vary within a match, the model predicts that all within-match wage variation is due to the evolution of beliefs about match quality. Specifically, m_τ is a martingale, so the model predicts that wages evolve as a martingale within a match.[12] This is the basis of a test of the matching model developed in Section 3.4.

The martingale property is common to other learning models, e.g. Farber and Gibbons (1996) and Gibbons et al. (2005). It implies that wage shocks are permanent and diminish with tenure,[13] and that the cross-sectional variance of wages increases with tenure (because $V_{\tau+1} > V_\tau$).[14]

[11] For instance, AKM, Goux and Maurin (1999), and ACK find negative correlations in France, ACK also find a similar result in Washington State, as do Gruetter and Lalive (2004) in Austrian data, and Barth and Dale-Olsen (2003) in Norwegian data.

[12] Due to the selection process that terminates a match if $m_\tau < \bar{m}_\tau$, the wage sequence observed by an econometrician is a submartingale.

[13] The definition of m_τ in equation (9) implies that shocks to beliefs about match quality (z and e_τ) are permanent. Within a match, these are the only shocks to wages, so wage shocks are permanent. Since each successive signal of match quality (shock) receives smaller weight in the Bayesian updating process, wage shocks diminish with tenure.

[14] Although the variance of beliefs (s_τ^2) declines with tenure because agents learn, the cross-sectional variance of m_τ increases with tenure. Specifically, all agents have common priors about match quality, hence the prior variance of m_τ is zero. As information accumulates,

The model also predicts that in a cross-section, workers with longer job tenure earn higher average wages than their counterparts with shorter tenure. This is consistent with stylized facts and numerous empirical findings, e.g., Mincer and Jovanovic (1981), Bartel and Borjas (1981), and many others.[15] The argument is as follows. First, larger values of θ and ψ are associated with higher wages and longer expected duration. Second, conditional on θ and ψ, better matches last longer and are associated with larger values of m_τ on average, and hence with higher wages. Third, because the reservation level of beliefs about match quality is monotone in tenure, the left tail of the wage distribution is increasingly truncated as tenure increases. All three effects operate in concert to induce a positive relationship between tenure and wages.

Finally, the model demonstrates that wages and match duration are jointly determined. This has important implications for our empirical specification. Since m_τ enters the equilibrium wage, but the match only continues as long as $m_\tau \geq \bar{m}_\tau$, the econometrician observes a truncated earnings distribution: earnings are only observed if $w_\tau \geq \gamma\mu + \theta + \psi + \gamma\bar{m}_\tau$. Since $\partial\bar{m}_\tau/\partial\theta < 0$ and $\partial\bar{m}_\tau/\partial\psi < 0$, this selection process induces a negative correlation between m_τ and the person and firm effects. Below, we develop a correction for the consequent selection bias.

3 Empirical Specification

Consider the following empirical counterpart to (26), where the first line coincides with the AKM model:

$$
\begin{aligned}
y_{ijt} &= \mu + x_{it}'\beta + \theta_i + \psi_j + \varepsilon_{ijt} && (35)\\
\varepsilon_{ijt} &= \gamma m_{ij\tau} + \nu_{ijt} && (36)
\end{aligned}
$$

where $i = 1, ..., N$ indexes workers and $j = 1, ..., J$ indexes firms; y_{ijt} is a measure of employment earnings; μ is the grand mean; x_{it} is a vector of observable time-varying covariates; β is a parameter vector; θ_i is the person effect; ψ_j is the firm effect of the firm j at which worker i was employed in t; and ε_{ijt} is a

m_τ converges to c, and V_τ increases from its prior value (zero) to its asymptotic value (σ_c^2) as $\tau \to \infty$.

[15] More recent research has focused on the causal link between job tenure and earnings growth using longitudinal data. Examples include Abraham and Farber (1987), Altonji and Shakotko (1987), Topel and Ward (1992), and Dostie (2005). The matching model implies that conditional on person and firm effects, all returns to tenure are due to accumulated knowledge about match quality. This accumulated knowledge is a form of (non-productive) match-specific human capital.

compound statistical error that consists of the posterior mean of beliefs about match quality $m_{ij\tau}$ times the worker's bargaining strength γ, plus classical measurement error ν_{ijt}. As in Section 2, τ indexes tenure. Equations (35) and (36) also introduce an additional index of calendar time, t.[16]

Some comments are in order. First, notice that (35) includes time-varying covariates x_{it}, whereas the equilibrium wage function (26) does not. The covariates admit variation over time in mean earnings ($\gamma\mu$ in equation (26)), and due to labor force experience and attachment.[17,18] Second, recall that the equilibrium wage function is specified in levels. Empirical models of earnings, however, are typically specified in logarithms.[19] For completeness, we estimate specifications on both log earnings and levels. We focus exclusively on the level specifications in the main text, since these correspond exactly to the matching model in Section 2. For interested readers, and to facilitate comparison with other research, we present estimates for the log specifications in appendicized tables.[20] Results for log and level specifications are qualitatively very similar, and they yield the same basic conclusions regarding the role of learning about match quality in earnings dynamics.

We further decompose the person effect θ_i into components observed and unobserved by the econometrician:

$$\theta_i = \alpha_i + u_i'\eta \tag{37}$$

where α_i is the unobserved component of the person effect; u_i is a vector

[16] The inclusion of time-varying covariates x_{it} in (35) necessitates the additional calendar time index t. Since tenure and calendar time are related by a simple function, I usually suppress one index.

[17] The vector of time-varying covariates x_{it} consists of year dummies, a quartic in experience (interacted with sex), and dummies for the number of quarters worked in the year (also interacted with sex).

[18] As noted in footnote 7, a more general model that includes deterministic human capital accumulation and publicly-observable stochastic aggregate productivity shocks yields an equlibrium wage that is additively separable in θ, ψ, m_τ, an experience effect, and time effects, like (35).

[19] We can think of the semi-log specification as a first-order approximation to the levels specification. If we rewrite $y_{ijt} = \mu\left(1 + x_{it}'\beta/\mu + \theta_i/\mu + \psi_j/\mu + \varepsilon_{ijt}/\mu\right)$, then $\ln y_{ijt} \approx \mu^* + x_{it}'\beta^* + \theta_i^* + \psi_j^* + \varepsilon_{ijt}^*$, where we have used the first-order Taylor series approximation $\ln(1+x) \approx x$ around $x = 0$, and where $\mu^* = \ln \mu$, $\beta^* = \beta/\mu$, $\theta_i^* = \theta_i/\mu$, $\psi_j^* = \psi_j/\mu$, and $\varepsilon_{ijt}^* = \varepsilon_{ijt}/\mu$.

[20] If we re-specify match output as $\exp\left(q_\tau\right)$, the logarithm of the equilibrium wage is additively separable in person and firm effects, m_τ, and s_τ^2 in the limiting case where $\left(1-\gamma\right)\left(1-\rho\right)U \to 0$. This coincides with the log-linear specification reported in the appendicized tables and figures, up to a tenure effect to capture the evolution of the variance of beliefs.

of time-invariant observable personal characteristics;[21] and η is a parameter vector.

Let N^* denote the total number of observations; q the number of time-varying covariates including the constant term; and p the number of time-invariant person characteristics. Rewriting (35) and (37) in matrix notation, we have

$$y = X\beta + U\eta + D\alpha + F\psi + \varepsilon \tag{38}$$

where y is the $N^* \times 1$ vector of earnings outcomes, X is the $N^* \times q$ matrix of time-varying covariates including the intercept; β is a $q \times 1$ parameter vector; U is the $N^* \times p$ matrix of time-invariant person characteristics; η is a $p \times 1$ parameter vector; D is the $N^* \times N$ design matrix of the unobserved component of the person effect; α is the $N \times 1$ vector of person effects; F is the $N^* \times J$ design matrix of the firm effects; ψ is the $J \times 1$ vector of firm effects; and ε is the $N^* \times 1$ vector of errors.

I consider two estimators of (38). The first is the ACK least squares estimator, that specifies α and ψ as fixed effects. The second estimator specifies α and ψ as random effects. We prefer the random effects estimator for a variety of reasons. First, as shown below, the least squares estimator is a limiting case of the random effects estimator. Second, the fixed effects estimator is less efficient because of the large number of person and firm effects to estimate. In the random effects model, in contrast, we only estimate a few parameters of the distribution of the random effects. Realized values of the random effects, the Best Linear Unbiased Predictors (BLUPs), are estimated in a subsequent step. Third, as discussed in Robinson (1991), BLUPs typically have better sampling properties than least squares estimates of fixed effects. Fourth, the random effects estimator allows us to estimate the within-match error covariance simultaneously via maximum likelihood; I use this estimate to test the learning hypothesis as described in Section 3.4. Finally, the random effects specification permits out-of-sample prediction of person and firm effects, which I use to validate the specification.

3.1 The Fixed Effect Estimator

The fixed model is completely specified by (38) and the stochastic assumptions:

$$E\left[\varepsilon | D, F, X, U\right] = 0 \tag{39}$$
$$E\left[\varepsilon\varepsilon'\right] = \sigma^2_\varepsilon I_{N^*} \tag{40}$$

[21] Characteristics in u_i include race (white/nonwhite/missing), education (5 categories), and an indicator for negative potential experience in the first period, all interacted with sex.

where I_{N^*} is the identity matrix of order N^*. Equation (39) embodies the so-called exogenous mobility assumption, since it rules out any systematic relationship between ε and employer identity (F). Under these assumptions, the least squares estimator is BLUE and solves the normal equations:

$$\begin{bmatrix} X'X & X'U & X'D & X'F \\ U'X & U'U & U'D & U'F \\ D'X & D'U & D'D & D'F \\ F'X & F'U & F'D & F'F \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\eta} \\ \hat{\alpha} \\ \hat{\psi} \end{bmatrix} = \begin{bmatrix} X'y \\ U'y \\ D'y \\ F'y \end{bmatrix}. \tag{41}$$

In our data, the cross product matrix on the left hand side of (41) is large enough to preclude estimation using standard software. I therefore compute least squares solutions $\hat{\beta}$, $\hat{\eta}$, $\hat{\alpha}$, and $\hat{\psi}$ using the ACK iterative conjugate gradient algorithm. The resulting estimates of α and ψ are not unique, since the design matrices D and F are not full rank; see ACK for a thorough discussion. I apply their grouping procedure to obtain unique estimates of α and ψ subject to the restriction that their overall mean, and their mean within each group of connected observations, are zero.[22] Beyond this normalization, the fixed effect estimator imposes no restrictions on the relationship between α, ψ, X and U.

3.2 The Random Effect Estimator

The random effect estimator treats β and η as fixed, and α and ψ as random effects. The model is completely specified by (38) and the stochastic assumptions:

$$E\left[\begin{array}{c|c} \alpha & \\ \psi & X, U \\ \varepsilon & \end{array}\right] = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad Var\left[\begin{array}{c|c} \alpha & \\ \psi & X, U \\ \varepsilon & \end{array}\right] = \begin{bmatrix} \sigma_\alpha^2 I_N & 0 & 0 \\ 0 & \sigma_\psi^2 I_J & 0 \\ 0 & 0 & R \end{bmatrix}. \tag{42}$$

It is worth noting that (42) imposes stronger assumptions on α and ψ than are required for the fixed effect estimator. However, these do not imply that D and F are orthogonal to X and U, which would usually be violated in economic data.

I estimate two alternate parameterizations of the error covariance R.[23] The simplest assumes spherical errors, $R = \sigma_\varepsilon^2 I_{N^*}$. I estimate this specification primarily for comparison with the fixed model. The second parameterization

[22] See Searle (1987) for a general discussion of connectedness. Here, firms are connected by a common employee and employees are connected by a common employer.

[23] Estimates of a variety of ARMA specifications of the error covariance are also available on request.

imposes no restrictions on the within-match error covariance other than symmetry and positive semi-definiteness.

3.2.1 REML Estimation of the Variance Parameters

I estimate the variance of the random effects $\left(\sigma_\alpha^2, \sigma_\psi^2\right)$ and R by Restricted Maximum Likelihood (REML).[24] REML is sometimes described as maximizing that part of likelihood that is invariant to the regression coefficients. Formally, it is maximum likelihood on linear combinations of y, chosen so that the linear combinations do not contain any of the coefficients. The linear combinations $k'y$ satisfy $k'\left(X\beta + U\eta\right) = 0 \ \forall \beta, \eta$, which implies $k' \left[\begin{array}{cc} X & U \end{array} \right] = 0$. Thus k' projects onto the space orthogonal to $\left[\begin{array}{cc} X & U \end{array} \right]$ and takes the form

$$k' = c' \left[I_{N^*} - \left[\begin{array}{cc} X & U \end{array} \right] \left(\left[\begin{array}{c} X' \\ U' \end{array} \right] \left[\begin{array}{cc} X & U \end{array} \right] \right)^{-} \left[\begin{array}{c} X' \\ U' \end{array} \right] \right] \equiv c' M_{XU} \qquad (43)$$

for arbitrary c', and where A^- denotes the generalized inverse of A. When $\left[\begin{array}{cc} X & U \end{array} \right]$ has rank $r \leq q + p$, there are $N^* - r$ linearly independent vectors k' satisfying (43). Define a matrix K' with rows k' satisfying (43) and full row rank $N^* - r$. REML estimation is maximum likelihood on $K'y$.[25]

The REML estimates of $\left(\sigma_\alpha^2, \sigma_\psi^2\right)$ and R have attractive properties. They are consistent, efficient, and asymptotically normal; see Jiang (1996). They are also invariant to the choice of K' and to the values of β and η.

3.2.2 Estimating the Coefficients and Realized Random Effects

The REML estimator does not directly estimate the coefficients β and η. These are obtained in a second step by solving the Henderson et al. (1959) mixed model equations for the BLUE of the coefficients and BLUPs of the random effects. If we define the matrix of variance components

$$G = \left[\begin{array}{cc} \sigma_\alpha^2 I_N & 0 \\ 0 & \sigma_\psi^2 I_J \end{array} \right], \qquad (44)$$

[24] REML estimators are widely used outside economics.

[25] I maximize the REML log-likelihood using the Average Information (AI) algorithm of Gilmour et al. (1995). The AI algorithm is a computationally convenient variant of Fisher scoring.

the mixed model equations are

$$\begin{bmatrix} \begin{bmatrix} X' \\ U' \end{bmatrix} R^{-1} \begin{bmatrix} X & U \end{bmatrix} & \begin{bmatrix} X' \\ U' \end{bmatrix} R^{-1} \begin{bmatrix} D & F \end{bmatrix} \\ \begin{bmatrix} D' \\ F' \end{bmatrix} R^{-1} \begin{bmatrix} X & U \end{bmatrix} & \begin{bmatrix} D' \\ F' \end{bmatrix} R^{-1} \begin{bmatrix} D & F \end{bmatrix} + G^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\eta} \\ \tilde{\alpha} \\ \tilde{\psi} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ U'R^{-1}y \\ D'R^{-1}y \\ F'R^{-1}y \end{bmatrix} \tag{45}$$

where $\tilde{\beta}, \tilde{\eta}, \tilde{\alpha}$, and $\tilde{\psi}$ denote solutions. The BLUPs $\tilde{\alpha}$ and $\tilde{\psi}$ are *best* in the sense of minimizing the mean square error of prediction, *linear* in y, and *unbiased* in the sense $E(\tilde{\alpha}) = E(\alpha)$ and $E(\tilde{\psi}) = E(\psi)$. See Robinson (1991) for details. In standard fashion, I solve (45) conditional on REML estimates of R and G.

The mixed model equations make clear the relationship between the fixed and random effect models. In particular, as $G \to \infty$ with $R = \sigma_\varepsilon^2 I_{N^*}$, the mixed model equations (45) converge to the normal equations (41). Thus the solutions $(\tilde{\beta}, \tilde{\eta}, \tilde{\alpha}, \tilde{\psi})$ converge to the least squares solutions $(\hat{\beta}, \hat{\eta}, \hat{\alpha}, \hat{\psi})$, so that the least squares estimator is a limiting case.

3.3 Correcting for Truncation of the Earnings Distribution

The matching model predicts that endogenous mobility due to learning about match quality truncates the earnings distribution. Specifically, earnings are only observed when $m_\tau \geq \bar{m}_\tau$ because the match terminates whenever m_τ falls below $\bar{m}_\tau$. This is potentially problematic for two reasons. First, the matching model implies that $\bar{m}_\tau$ depends on worker and firm quality, which imparts selection bias in the wage equation. Second, our test of the learning hypothesis (Section 3.4) relies on estimates of $Var[m_\tau]$ and $Cov[m_\tau, m_{\tau'}]$, which are subject to truncation bias.

If we iterate forward on the definition of $\bar{m}_\tau$ in (30) we obtain

$$\bar{m}_\tau = -[\mu A_\tau + B_\tau] - [a - U(1-\rho)] A_\tau - bA_\tau \tag{46}$$

where

$$\begin{aligned} A_\tau &= 1 + \sum_{s=1}^{\infty} [\rho(1-\delta)]^s \int_{\bar{m}_{\tau+1}}^{\infty} \int_{\bar{m}_{\tau+2}}^{\infty} \cdots \int_{\bar{m}_{\tau+s}}^{\infty} dF_{\tau+s} \cdots dF_{\tau+2} d\bar{F}_{\tau+1} \\ B_\tau &= \sum_{s=1}^{\infty} [\rho(1-\delta)]^s \int_{\bar{m}_{\tau+1}}^{\infty} \int_{\bar{m}_{\tau+2}}^{\infty} \cdots \int_{\bar{m}_{\tau+s}}^{\infty} m_{\tau+s} dF_{\tau+s} \cdots dF_{\tau+2} d\bar{F}_{\tau+1} \end{aligned}$$

where $F_\tau = F\left(m_\tau | m_{\tau-1}, s^2_{\tau-1}\right)$ and $\bar{F}_\tau = F\left(m_\tau | \bar{m}_{\tau-1}, s^2_{\tau-1}\right)$. Reintroducing the person and firm subscripts, I approximate (46) by

$$\bar{m}_{ij\tau} = -\mu_\tau - \zeta_{i\tau} - \xi_{j\tau}. \tag{47}$$

Recall $m_{ij\tau} \sim N(0, V_\tau)$. Under the approximation (47), the probability that the match between worker i and firm j persists to tenure τ is

$$\Pr\left(m_{ij\tau} \geq \bar{m}_{ij\tau}\right) = 1 - \Phi\left(\frac{-\mu_\tau - \zeta_{i\tau} - \xi_{j\tau}}{V_\tau^{1/2}}\right) = \Phi\left(\frac{\mu_\tau + \zeta_{i\tau} + \xi_{j\tau}}{V_\tau^{1/2}}\right) \tag{48}$$

where Φ is the standard normal CDF. The conditional expectation of observed earnings is therefore

$$\begin{aligned} E\left[y_{ijt} | m_{ij\tau} \geq \bar{m}_{ij\tau}\right] &= \mu + x'_{it}\beta + \theta_i + \psi_j + \gamma V_\tau^{1/2} \frac{\phi\left(\frac{\mu_\tau + \zeta_{i\tau} + \xi_{j\tau}}{V_\tau^{1/2}}\right)}{\Phi\left(\frac{\mu_\tau + \zeta_{i\tau} + \xi_{j\tau}}{V_\tau^{1/2}}\right)} \\ &= \mu + x'_{it}\beta + \theta_i + \psi_j + \gamma V_\tau^{1/2} \lambda_{ij\tau} \end{aligned} \tag{49}$$

where $\lambda_{ij\tau}$ is the familiar Inverse Mills' Ratio.

I perform a two-step truncation correction based on (48) and (49). At each τ, I estimate a probit model of the probability that the match continues for one more period. The probits include random person and firm effects,[26] and are estimated by Average Information REML applied to the method of Schall (1991).[27] With estimates of the realized random effects $\tilde{\zeta}_{it}$ and $\tilde{\xi}_{j\tau}$ in hand, I construct an estimate of the Inverse Mills' Ratio, $\tilde{\lambda}_{ij\tau}$, and include it as an additional covariate in the earnings equation. This corrects the estimated coefficients, person effects, and firm effects for selection due to endogenous mobility. As described below, I also use $\tilde{\lambda}_{ij\tau}$ to correct the estimated within-match error covariance for truncation bias.

3.4 The Learning Hypothesis

Learning about match quality implies a specific structure for the within-match error covariance, W. Consider the balanced data case, where all matches last

[26] Identification of the random effects required pooling of some probit equations across tenure levels.

[27] Schall (1991) extends standard estimation methods for generalized linear models to the random effects case. It is based on REML estimation of a linearized link function (in this case, Φ).

$\bar{\tau}$ periods so that $R = I_M \otimes W$ and W is $\bar{\tau} \times \bar{\tau}$. The compound error due to learning and measurement error, (36), implies

$$W = \begin{bmatrix} \gamma^2 V_1 + \sigma_\nu^2 & \gamma^2 V_1 & \gamma^2 V_1 & \cdots & \gamma^2 V_1 \\ \gamma^2 V_1 & \gamma^2 V_2 + \sigma_\nu^2 & \gamma^2 V_2 & \cdots & \gamma^2 V_2 \\ \gamma^2 V_1 & \gamma^2 V_2 & \gamma^2 V_3 + \sigma_\nu^2 & \cdots & \gamma^2 V_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma^2 V_1 & \gamma^2 V_2 & \gamma^2 V_3 & \cdots & \gamma^2 V_{\bar{\tau}} + \sigma_\nu^2 \end{bmatrix} \tag{50}$$

where σ_ν^2 is the variance of measurement error.[28] Note that (50) does not account for truncation of the error distribution that arises because of endogenous mobility due to learning about match quality. Given a suitable correction for truncation bias, however, we can test whether the empirical covariance coincides with (50). Furthermore, since the structural parameters σ_c^2, σ_z^2, and σ_e^2 enter into each V_τ, (recall its definition in eq. (10)), they can be recovered from an estimate of W.

I test the learning hypothesis and recover the structural parameters and σ_ν^2 in two steps. First, I estimate the within-match error covariance. For the fixed effect case, I use the within-match sample covariance of the residuals. For the random effect case, I use the unrestricted REML estimate of the within-match error covariance. Given the probit models of Section 3.3, I correct the diagonal elements of the covariance matrix for truncation bias as in Heckman (1979), and the off-diagonal elements as in Alexander et al. (1987).[29] Second, following Abowd and Card (1989) and Farber and Gibbons (1996), I fit the learning structure (50) to the estimated error covariance by minimum distance.[30] This yields estimates of the structural parameters up to the factor γ^2. Finally, I test the overidentifying restrictions implied by (50) using the Newey (1985) test statistic.[31]

---

[28] In the unbalanced data case, where the match between worker i and firm j lasts τ_{ij} periods, R is a block-diagonal matrix with $\tau_{ij} \times \tau_{ij}$ diagonal blocks composed of the first τ_{ij} rows and columns of (50).

[29] The Alexander et al. (1987) correction accounts for truncation in both covarying variables

[30] Optimal minimum distance estimation, as in Hansen (1982) and Chamberlain (1984), proved infeasible because our estimates of W were poorly conditioned and did not invert. I use a diagonal weight matrix instead, with elements equal to the logarithm of the number of observations contributing to that element of W. This gives greater weight to more precisely estimated elements of W. Weighting by a diagonal matrix of variances yields similar results, as does equally weighted minimum distance.

[31] The Newey (1985) test statistic does not require inverting the variance of the moment conditions.

4 Data

Identifying the person and firm effects requires repeated observations on both workers and firms. I use data from the US Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program database. The LEHD data span 48 states. In this paper, I use data from two states. Their identity is confidential.

The LEHD data are administrative, constructed from quarterly Unemployment Insurance (UI) system wage reports. The characteristics of the UI wage data vary slightly from state to state. However the Bureau of Labor Statistics (1997, p. 42) claims that UI coverage is "broad and basically comparable from state to state" and that "over 96 percent of total wage and salary civilian jobs" were covered in 1994. See Abowd et al. (2009) for details. With the UI wage records as its frame, the LEHD data comprise the universe of employers required to file UI system wage reports — that is, all employment potentially covered by the UI system in participating states.

The UI wage records themselves contain only very limited information: worker, firm, and time identifiers, and employment earnings. Reported earnings include gross wages and salary, bonuses, stock options, tips and gratuities, and the value of meals and lodging when these are supplied (Bureau of Labor Statistics (1997, p. 44)). The LEHD database integrates the UI wage records with internal Census Bureau data to add demographic and firm characteristics, including sex, race, date of birth, industry, and geography.

I aggregate the underlying quarterly data to the annual level for estimation. The full sample consists of over 49 million annualized employment records on more than 9 million full-time workers between 25 and 65 years of age who were employed at nearly 575,000 private-sector non-agricultural firms between 1990 and 1999.

Missing data items are multiply-imputed: three imputed values are generated for each missing item. The result is three versions of the database ("implicates"), each of which contains a different set of imputed values. See the Data Appendix for further details on sample construction, variable creation, and missing data imputation.

The computational efficiency of the ACK algorithm allows me to estimate the fixed effect specification on the full sample. Computational demands of the random effect estimator, however, necessitate estimating it on a subsample. Sampling from these data is nontrivial because the sample must be sufficiently connected to precisely estimate the person and firm effects.[32] I therefore draw a

[32] Identifying the effects requires multiple observations on employees of most firms and mo-

TABLE 1
PROPERTIES OF CONNECTED GROUPS OF WORKERS AND FIRMS

	Full Sample[a]	Dense Sample 1[b]	Simple Random Sample[c]
Number of Groups	84,708	1,140	9,457
Number of Workers	9,271,766	49,425	49,200
Number of Firms	573,237	27,421	40,064
Number of Worker-Firm Matches	15,305,508	92,539	93,182
Number of Matches in Smallest Group	1	5	1
Percent of Matches in:			
Largest Group	99.06	67.25	59.37
Second Largest Group	0.0006	24.70	20.30
Third Largest Group	0.0003	0.04	0.06
Groups containing 5 or more matches	99.21	100	84.44
Groups containing only 1 match	0.35	0	5.50

[a] Results combined across three completed data implicates.

[b] One percent dense random samples of workers employed in 1997. Results are combined across three implicates.

[c] One percent simple random sample of workers employed in 1997. Results are for one completed data implicate.

one percent subsample using the dense sampling algorithm described in Woodcock (2007). This sampling procedure ensures that each worker is connected to at least five others by a common employer, but is otherwise equivalent to a simple random sample of individuals employed in a reference year (1997). That is, all individuals employed in 1997 have an equal probability of being sampled. A second (disjoint) one percent dense random subsample is drawn for model validation. In what follows, these are called Dense Sample 1 and Dense Sample 2, respectively. All random effect estimates are computed on Dense Sample 1. Tables 1 and 2 report characteristics of the samples; see the Data Appendix for more information.

Estimating the within-match error covariance requires a measure of job tenure. All employment spells active in the database's first quarter are presumed left-censored. Since job tenure is unknown for left-censored spells, they are excluded from estimates of W and our tests of the learning hypothesis. Those estimates that do not require a measure of job tenure include the left-censored spells.

---

bility of workers between firms. A small simple random sample of individuals, for example, is usually not sufficiently connected to estimate the person and firm effects precisely.

TABLE 2
SUMMARY STATISTICS
Estimates Combined Across 3 Completed Data Implicates

	Full Sample		Dense Sample 1		Simple Random Sample	
Variable	Mean[a]	Std. Dev[b]	Mean[a]	Std. Dev[b]	Mean[a]	Std. Dev[b]
Demographic Characteristics						
Male (Proportion)	0.56	0.50	0.56	0.50	0.57	0.50
Age (Years)	40.6	10.2	40.4	9.5	40.3	9.5
Men						
Nonwhite (Proportion)	0.21	0.57	0.21	0.57	0.21	0.57
Race Missing (Proportion)	0.04	0.25	0.03	0.24	0.03	0.24
Less than high school (Proportion)	0.12	0.45	0.11	0.43	0.11	0.42
High school (Proportion)	0.30	0.67	0.29	0.66	0.30	0.66
Some college (Proportion)	0.23	0.60	0.23	0.60	0.23	0.59
Associate or Bachelor Degree (Proportion)	0.25	0.62	0.26	0.62	0.26	0.62
Graduate or Professional Degree (Proportion)	0.10	0.42	0.11	0.43	0.11	0.42
Women						
Nonwhite (Proportion)	0.24	0.69	0.24	0.70	0.24	0.60
Race Missing (Proportion)	0.02	0.22	0.02	0.22	0.02	-0.01
Less than high school (Proportion)	0.09	0.45	0.09	0.43	0.09	0.28
High school (Proportion)	0.31	0.78	0.30	0.77	0.30	0.39
Some college (Proportion)	0.25	0.71	0.25	0.71	0.25	0.31
Associate or Bachelor Degree (Proportion)	0.26	0.72	0.28	0.75	0.27	0.42
Graduate or Professional Degree (Proportion)	0.08	0.42	0.09	0.44	0.09	0.31
Work History Characteristics						
Tenure (Years)	4.5	3.5	4.9	3.6	4.8	3.6
Job is Left Censored (Proportion)	0.33	0.47	0.36	0.48	0.35	0.48
Real Annualized Earnings (1990 Dollars)	53755	50804	57209	51196	56483	50074

TABLE 2, Continued

	Full Sample		Dense Sample 1		Simple Random Sample	
Variable	Mean[a]	Std. Dev[b]	Mean[a]	Std. Dev[b]	Mean[a]	Std. Dev[b]
Men						
Labor Market Experience (Years)	11.8	13.1	11.7	12.6	11.7	12.6
Initial Experience <0 (Proportion)	0.02	0.20	0.02	0.20	0.02	0.20
Worked 0 Full Quarters in Calendar Year (Proportion)	0.08	0.36	0.06	0.32	0.06	0.32
Worker 1 Full Quarter in Calendar Year (Proportion)	0.15	0.49	0.11	0.44	0.12	0.44
Worker 2 Full Quarters in Calendar Year (Proportion)	0.13	0.47	0.12	0.45	0.12	0.45
Worker 3 Full Quarters in Calendar Year (Proportion)	0.14	0.48	0.14	0.47	0.13	0.47
Worker 4 Full Quarters in Calendar Year (Proportion)	0.50	0.91	0.57	0.96	0.56	0.97
Women						
Labor Market Experience (Years)	9.5	13.0	9.3	12.6	9.2	12.5
Initial Experience <0 (Proportion)	0.02	0.23	0.02	0.23	0.02	0.22
Worked 0 Full Quarters in Calendar Year (Proportion)	0.07	0.39	0.05	0.35	0.06	0.36
Worker 1 Full Quarter in Calendar Year (Proportion)	0.14	0.54	0.11	0.49	0.11	0.50
Worker 2 Full Quarters in Calendar Year (Proportion)	0.13	0.53	0.12	0.50	0.12	0.51
Worker 3 Full Quarters in Calendar Year (Proportion)	0.14	0.55	0.13	0.53	0.13	0.53
Worker 4 Full Quarters in Calendar Year (Proportion)	0.52	0.96	0.59	1.00	0.59	1.01
Number of Observations	49,281,533		357,725		357,009	
Number of Workers	9,271,766		49,425		49,200	
Number of Firms	573,237		27,421		40,064	
Number of Worker-Firm Matches	15,305,508		92,539		93,182	

[a] Means are computed on each completed data implicate for each sample. The reported value is the arithmetic mean of the means computed on each implicate.

[b] Square root of the total variance, as defined in Rubin (1987).

5 Results

I estimate the specifications described in Section 3 on each completed data implicate, and combine parameter estimates across the implicates using standard formulae in Rubin (1987). Estimates of the coefficients β and η are available from the author upon request. Their values are reasonable and vary little across specifications.

5.1 Parameter Estimates and Model Fit

Tables 3 and 4 present the estimated variance components and a summary of model fit for the three specifications of interest: the fixed effect estimator, the random effect estimator with spherical errors ($R = \sigma_{\varepsilon}^2 I_{N^*}$), and the random effect specification with unrestricted within-match error covariance. Estimates in Table 3 include left-censored spells,[33] while those in Table 4 exclude them. The two sets of estimates are very similar.

The random effect model with spherical errors obtains the best fit by in-sample measures that penalize for parameterization (AIC, BIC). To assess the out-of-sample performance of the random effect models, I solve the mixed model equations (45) on Dense Sample 2, using the variance components $\tilde{G}$ and error covariance $\tilde{R}$ estimated on Dense Sample 1, and report the variance of prediction errors.[34] The two random effect specifications perform similarly in this regard, and the variance of prediction errors is close to the estimated error variance. I also test the random effect specifications against the fixed effect alternative using the usual Hausman (1978) test. Unfortunately the test statistics were negative and hence the tests inconclusive.

The estimated variance components have a straightforward interpretation. I scale earnings to have unit variance so that, conditional on all other effects, a one standard deviation increase in α_i increases real annualized earnings by σ_α standard deviations. Similarly, a one standard deviation increase in ψ_j increases real annualized earnings by σ_ψ standard deviations.

The estimated variance of the person effect is much larger than the variance of the firm effect in all specifications. Evidently, individual heterogeneity generates much more earnings dispersion than firm heterogeneity. This is consistent with the findings of AKM and others. In Tables 3 and 4, the fixed

[33] Table 3 estimates of the mixed model with W unrestricted are based on multiply-imputed tenure for left-censored spells.

[34] For models with $R = \sigma_{\varepsilon}^2 I_{N^*}$, prediction error is the estimated residual. For the model with W unrestricted, prediction error is the difference between the residual and its conditional expectation given the other within-match residuals under multivariate normality.

TABLE 3
ESTIMATED VARIANCE COMPONENTS AND SUMMARY OF MODEL FIT
Unit Variance Scale, Combined Results From 3 Completed Data Implicates

	Fixed Effect Estimator		Random Effects With Spherical Error		Random Effects With Unrestricted Within-Match Error Covariance	
	Estimate[a,e]	Std Error[b]	Estimate[a]	Std Error[b]	Estimate[a,f]	Std Error[b]
No Correction for Truncation						
Variance of person effect (σ^2_α)	0.720	(0.003)	0.551	(0.004)	0.402	(0.002)
Variance of firm effect (σ^2_ψ)	0.062	(0.000)	0.100	(0.003)	0.043	(0.007)
Residual variance (σ^2_ε)	0.160	(0.000)	0.151	(0.001)	n/a	n/a
AIC[c]	-1.44	(0.000)	-1.89	(0.006)	-1.46	(0.023)
BIC[c]	1.67	(0.000)	-1.89	(0.006)	-1.46	(0.023)
V(out-of-sample pred. error)[c,d]			0.131	(0.001)	0.131	(0.001)
Corrected for Truncation						
Variance of person effect (σ^2_α)			0.551	(0.036)	0.401	(0.002)
Variance of firm effect (σ^2_ψ)			0.099	(0.003)	0.043	(0.006)
Residual variance (σ^2_ε)			0.151	(0.001)	n/a	n/a
Truncation Correction (β_λ)			0.045	(0.005)	0.069	(0.005)
AIC[c]			-1.89	(0.006)	-1.46	(0.014)
BIC[c]			-1.89	(0.006)	-1.46	(0.014)
V(out-of-sample pred. error)[c,d]			0.131	(0.001)	0.131	(0.001)
Number of Observations	49,281,533	(9103)	357,725	(2363)	357,725	(2363)
Number of Workers	9,271,766	(710)	49,425	(150)	49,425	(150)
Number of Firms	573,237	(118)	27,421	(13)	27,421	(13)
Number of Matches	15,305,508	(3196)	92,539	(470)	92,539	(470)

[a] Mean parameter estimate in three completed data implicates.

[b] Square root of total variance of parameter estimate over three completed data implicates, as defined in Rubin (1987).

[c] Value in column labeled "Std Error" is the between-implicate standard deviation.

[d] Variance of out-of-sample prediction errors in Dense Sample 2.

[e] Sample variance of estimated person and firm effects, averaged over three completed data implicates.

[f] Estimates based on multiply-imputed tenure for left-censored spells.

TABLE 4
ESTIMATED VARIANCE COMPONENTS AND SUMMARY OF MODEL FIT
Unit Variance Scale, Combined Results From 3 Completed Data Implicates
Left-Censored Spells Excluded

	Fixed Effect Estimator		Random Effects With Spherical Error		Random Effects With Unrestricted Within-Match Error Covariance	
	Estimate[a,e]	Std Error[b]	Estimate[a]	Std Error[b]	Estimate[a]	Std Error[b]
No Correction for Truncation						
Variance of person effect (σ^2_α)	0.700	(0.003)	0.567	(0.005)	0.461	(0.010)
Variance of firm effect (σ^2_ψ)	0.068	(0.000)	0.103	(0.004)	0.047	(0.002)
Residual variance (σ^2_ε)	0.161	(0.000)	0.160	(0.001)	n/a	n/a
AIC[c]	-1.33	(0.000)	-1.83	(0.003)	-1.76	(0.028)
BIC[c]	2.41	(0.001)	-1.83	(0.003)	-1.75	(0.028)
V(out-of-sample pred. error)[c,d]			0.132	(0.001)	0.139	(0.002)
Corrected for Truncation						
Variance of person effect (σ^2_α)			0.567	(0.039)	0.461	(0.010)
Variance of firm effect (σ^2_ψ)			0.102	(0.004)	0.046	(0.002)
Residual variance (σ^2_ε)			0.160	(0.001)	n/a	n/a
Truncation Correction (β_λ)			0.050	(0.006)	0.070	(0.005)
AIC[c]			-3.01	(0.018)	-1.75	(0.028)
BIC[c]			-3.01	(0.018)	-1.75	(0.028)
V(out-of-sample pred. error)[c,d]			0.132	(0.001)	0.139	(0.002)
Number of Observations	32,800,936	(7217)	228,386	(2018)	228,386	(2018)
Number of Workers	7,577,051	(2855)	39,816	(168)	39,816	(168)
Number of Firms	544,254	(177)	24,624	(22)	24,624	(22)
Number of Matches	12,289,989	(694)	73,307	(475)	73,307	(475)

[a] Mean parameter estimate in three completed data implicates.

[b] Square root of total variance of parameter estimate over three completed data implicates, as defined in Rubin (1987).

[c] Value in column labeled "Std Error" is the between-implicate standard deviation.

[d] Variance of out-of-sample prediction errors in Dense Sample 2.

[e] Sample variance of estimated person and firm effects, averaged over three completed data implicates.

effect specification yields the largest estimate of σ_α^2 (around 0.7), but among the smallest estimates of σ_ψ^2 (0.07). The random effect model with spherical errors yields a slightly smaller estimate of σ_α^2 (about 0.55), and a slightly larger estimate of σ_ψ^2. Relaxing the spherical errors assumption in favor of the unrestricted within-match error covariance reduces the estimates of both σ_α^2 and σ_ψ^2, to 0.4 and 0.04, respectively. This reduction in variance is consistent with the matching model, which predicts that earnings covariation within a match is stronger (relative to covariation between matches) than implied by covariation due to person effects, firm effects, and spherical errors alone, because of covariation between the $m_{ij\tau}$. Specifications with spherical errors attribute all within-match covariation to the person and firm effects, leading us to over-estimate their variance.[35]

The truncation correction has only a minor impact on the estimated variance components. Evidently selection due to learning about match quality imparts only limited bias on the estimated variance of person and firm effects. This implies that the relationship between the threshold level of beliefs about match quality, $\bar{m}_\tau$, and worker/firm quality is relatively weak; i.e., the derivatives (31) are close to zero. We continue to prefer the truncation-corrected estimates, however, and hence focus on these in the remainder. Furthermore, we shall see below that the truncation correction is important for estimating the within-match error covariance.

Table 5 presents correlations among the estimated effects. They are qualitatively similar across specifications. We use these to decompose earnings variation into the proportion attributed to each effect, by noting that

$$\begin{aligned} Var[y] &= Cov[y, X\beta + \theta + \psi + \varepsilon] \\ &= Cov[y, X\beta] + Cov[y, \theta] + Cov[y, \psi] + Cov[y, \varepsilon] = 1. \end{aligned}$$

For the truncation-corrected estimates of the random effect model with unrestricted W, this decomposition attributes 51 percent of earnings variation to θ_i, 6.3 percent to ψ_j, and 7.8 percent to time-varying observables, leaving 34.9 percent unexplained. After estimating the remaining structural parameters in Section 5.2, we decompose this unexplained variation into components attributable to learning about match quality and measurement error.

[35] Woodcock (2007) reports a similar result based on a specification with random person, firm, and match effects. That specification is the special case of (35) that arises when match quality is observed by workers and firms. Like our specification with W unrestricted, it relaxes the within-match covariance of wages relative to the person and firm effects model with spherical errors. Woodcock (2007) derives the bias caused by omitted match effects, and shows that it inflates estimates of σ_α^2 and σ_ψ^2.

TABLE 5
CORRELATIONS AMONG ESTIMATED EFFECTS
Combined Results From 3 Completed Data Implicates

No Correction for Truncation

Fixed Effect Estimator	y	θ	α	Uη	ψ	Xβ
Log Earnings (y)	1					
Total Person Effect (θ)	0.76	1				
Unobserved Component (α)	0.73	0.94	1			
Observed Component (Uη)	0.22	0.35	0.00	1		
Total Firm Effect (ψ)	0.30	0.02	0.01	0.04	1	
Time-Varying Covariates (Xβ)	0.20	-0.29	-0.23	-0.22	0.06	1
Random Effects With Spherical Error						
Log Earnings (y)	1					
Total Person Effect (θ)	0.85	1				
Unobserved Component (α)	0.81	0.94	1			
Observed Component (Uη)	0.25	0.34	0.01	1		
Total Firm Effect (ψ)	0.40	0.17	0.16	0.07	1	
Time-Varying Covariates (Xβ)	0.26	-0.06	-0.01	-0.14	0.02	1
Random Effects With Unrestricted Error Covariance						
Log Earnings (y)	1					
Total Person Effect (θ)	0.81	1				
Unobserved Component (α)	0.78	0.91	1			
Observed Component (Uη)	0.24	0.42	0.01	1		
Total Firm Effect (ψ)	0.41	0.26	0.25	0.07	1	
Time-Varying Covariates (Xβ)	0.26	-0.08	-0.01	-0.16	0.04	1

Corrected For Truncation

Random Effects With Spherical Error	y	θ	α	Uη	ψ	Xβ
Log Earnings (y)	1					
Total Person Effect (θ)	0.85	1				
Unobserved Component (α)	0.81	0.94	1			
Observed Component (Uη)	0.25	0.34	0.01	1		
Total Firm Effect (ψ)	0.40	0.17	0.16	0.07	1	
Time-Varying Covariates (Xβ)	0.26	-0.06	-0.01	-0.14	0.02	1
Truncation Correction ($\beta_\lambda\lambda$)	-0.01	-0.03	-0.03	-0.03	-0.10	-0.04
Random Effects With Unrestricted Error Covariance						
Log Earnings (y)	1					
Total Person Effect (θ)	0.81	1				
Unobserved Component (α)	0.78	0.91	1			
Observed Component (Uη)	0.24	0.42	0.01	1		
Total Firm Effect (ψ)	0.41	0.26	0.25	0.07	1	
Time-Varying Covariates (Xβ)	0.26	-0.07	-0.01	-0.15	0.04	1
Truncation Correction ($\beta_\lambda\lambda$)	-0.01	-0.03	-0.03	-0.04	-0.15	-0.03

Recall that the matching model predicts a negative correlation between θ_i and ψ_j. The fixed effect estimator finds a near-zero correlation (0.02). This is similar to Abowd et al.'s (2003) estimate, based on seven states from the LEHD data, and ACK, based on Washington State data. In contrast, the random effect specifications all exhibit a more substantial positive correlation between θ_i and ψ_j: 0.17 for the case of spherical errors, and 0.26 when the specification is relaxed to allow an unrestricted within-match error covariance.[36]

The sign of the correlation between θ_i and ψ_j contradicts a prediction of the matching model. However, Figures 1 and 2, which plot a nonparametric estimate of the regression of θ_i on ψ_j, suggest that focusing solely on correlations may be misleading: the relationship between person and firm effects is not monotone. There is a systematic but weak positive association between θ_i and ψ_j in the neighborhood of $\theta_i = \psi_j = 0$. However there is evidence of mismatch among more extreme values. A richer model of how workers sort across firms may be necessary to explain this non-monotonicity.

We note that correlations involving the truncation correction term $(\beta_\lambda \lambda_{ij\tau})$ are consistent with the matching model. Recall that $\beta_\lambda \lambda_{ij\tau}$ is an estimate of $\gamma E\left[m_\tau | m_\tau > \bar{m}_\tau\right]$. Since this is increasing in $\bar{m}_\tau$, and recalling (31) and lemma 6, the matching model predicts $\partial \beta_\lambda \lambda_{ij\tau} / \partial \theta < 0$ and $\partial \beta_\lambda \lambda_{ij\tau} / \partial \psi < 0$. The data support this prediction: correlations between $\beta_\lambda \lambda_{ij\tau}$ and $\left(\theta_i, \psi_j\right)$ are negative in both random effect specifications. However they are small in absolute value, providing further evidence of a weak relationship between $\bar{m}_\tau$ and worker/firm quality.

5.2 Testing the Learning Hypothesis

Table 6 presents estimates of the within-match error covariance. Not surprisingly, the truncation correction has a more substantive impact here than above. Even if $\bar{m}_\tau$ varies little across workers and firms (for a given τ), learning about match quality still truncates the error distribution because agents will terminate matches when they learn that match quality is low. Absent a correction, this truncation biases the estimated error variance-covariance.

Estimates from the random effect specification exhibit the basic properties of the learning structure in (50): in each column, the diagonal elements are larger than the off-diagonal elements (due to measurement error), elements increase in magnitude from left to right within each row (because $V_{\tau+1} > V_\tau$),

[36] The positive correlation between θ_i and ψ_j would seem to violate (42). However, Woodcock (2007) derives the covariance between $\tilde{\theta}_i$ and $\tilde{\psi}_j$ and shows it is positive via both frequentist and Bayesian derivations. From a Bayesian perspective, this reflects uncertainty over the attribution of wage variation to person and firm effects, conditional on X, U.

Figure 1

Estimated Regression of Person Effect (Theta) on Firm Effect (Psi)
Random Effects With Spherical Errors, Corrected for Truncation

Figure 2

Estimated Regression of Person Effect (Theta) on Firm Effect (Psi)
Random Effects With W Unrestricted, Corrected for Truncation

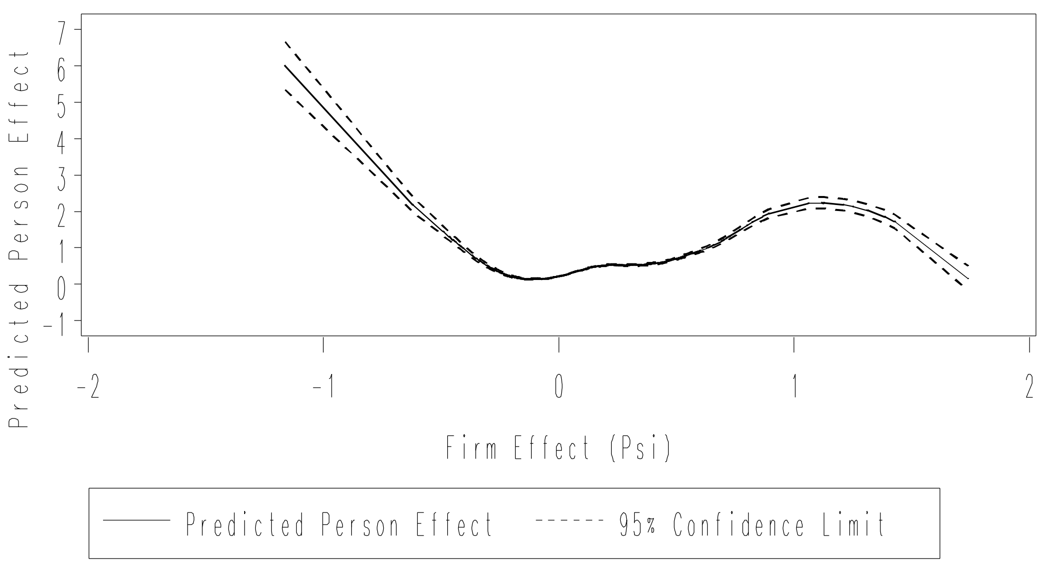


TABLE 6
WITHIN-MATCH ERROR COVARIANCE
Left-Censored Spells Excluded, Results Combined from 3 Implicates

Fixed Effect Estimator

Tenure	1	2	3	4	5	6	7	8	9	10
1	.157									
2	.034	.133								
3	.002	.041	.149							
4	-.021	.009	.039	.161						
5	-.041	-.016	.007	.043	.176					
6	-.061	-.038	-.016	.011	.053	.198				
7	-.079	-.055	-.035	-.009	.020	.065	.221			
8	-.095	-.070	-.051	-.026	-.002	.033	.081	.249		
9	-.113	-.086	-.066	-.041	-.018	.011	.045	.102	.288	
10	-.140	-.106	-.083	-.052	-.033	-.010	.021	.064	.126	.338

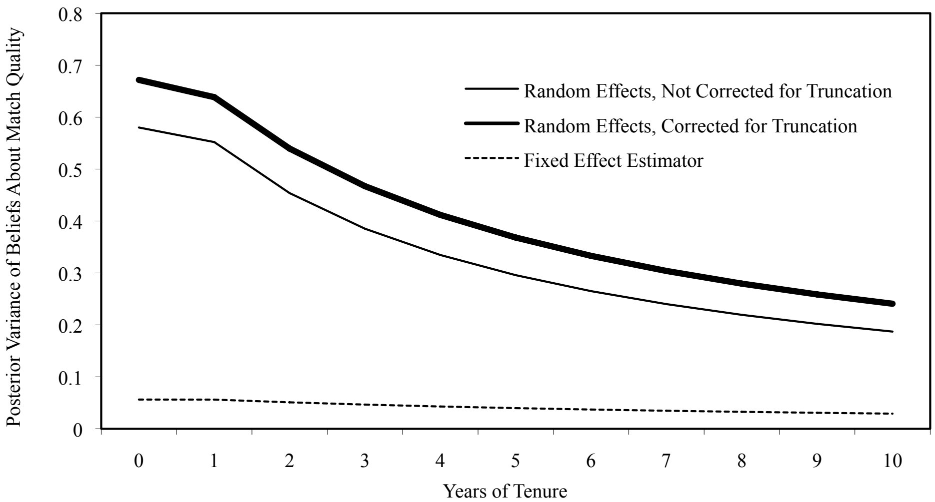
Random Effect Estimator

Tenure	1	2	3	4	5	6	7	8	9	10
1	.183									
2	.069	.164								
3	.060	.116	.195							
4	.060	.115	.167	.289						
5	.054	.099	.152	.226	.302					
6	.039	.109	.152	.221	.255	.369				
7	.056	.116	.163	.235	.264	.327	.443			
8	.070	.095	.146	.227	.266	.319	.380	.495		
9	.075	.108	.153	.227	.265	.309	.370	.435	.551	
10	.064	.113	.155	.224	.266	.312	.375	.430	.485	.612

Random Effect Estimator, Corrected for Truncation

Tenure	1	2	3	4	5	6	7	8	9	10
1	.240									
2	.074	.221								
3	.062	.134	.246							
4	.061	.126	.189	.342						
5	.055	.105	.168	.254	.355					
6	.039	.116	.166	.243	.284	.424				
7	.057	.124	.176	.255	.290	.361	.496			
8	.072	.098	.155	.244	.289	.347	.415	.551		
9	.077	.113	.162	.242	.285	.333	.399	.472	.606	
10	.066	.119	.164	.237	.285	.335	.402	.463	.523	.669

Figure 3
Estimated Sequence of Belief Variances



and off-diagonal elements within each column are approximately equal. Although the first two properties are also reflected in the fixed effect estimates, the latter is clearly not. In that model, there is a consistent decline in auto-covariances as one moves from lower to higher orders.

I fit the learning structure (50) to the fixed and random effect estimates of W by minimum distance. Table 7 reports the results on the scale of the data, i.e., for $\gamma = 1$.[37]

Parameter estimates suggest that learning about match quality is quite slow. The estimated variance of the initial signal (σ_z^2) is very large in all specifications, implying that it conveys almost no information. Similarly, production outcomes are noisy ($\sigma_e^2 = 3.48$ after correcting for truncation) and hence convey limited information about match quality. This is confirmed in Figure 3, which plots the variance of beliefs about match quality, s_τ^2. The truncation-corrected estimates imply that the initial signal only moderately reduces the variance of beliefs from its prior value of 0.67 to 0.63. After observing one period of production, it falls modestly to 0.54. It falls further to 0.37 by the fifth year, and to 0.24 by the tenth year. Hence learning about match quality takes time, but does slowly reveal the quality of the match. As for the test of over-identifying restrictions, we fail to reject the learning hypothesis at the 5% level of significance on the uncorrected error covariance,

[37] Recall that σ_c^2, σ_z^2, and σ_e^2 are only identified up to γ^2. Estimates can be re-scaled for any other $0 < \gamma < 1$: the re-scaled parameter is $\sigma_*^2 = \sigma^2/\gamma^2$.

TABLE 7

MINIMUM DISTANCE ESTIMATES OF STRUCTURAL PARAMETERS

Combined Results From 3 Completed Data Implicates, Left-Censored Spells Excluded

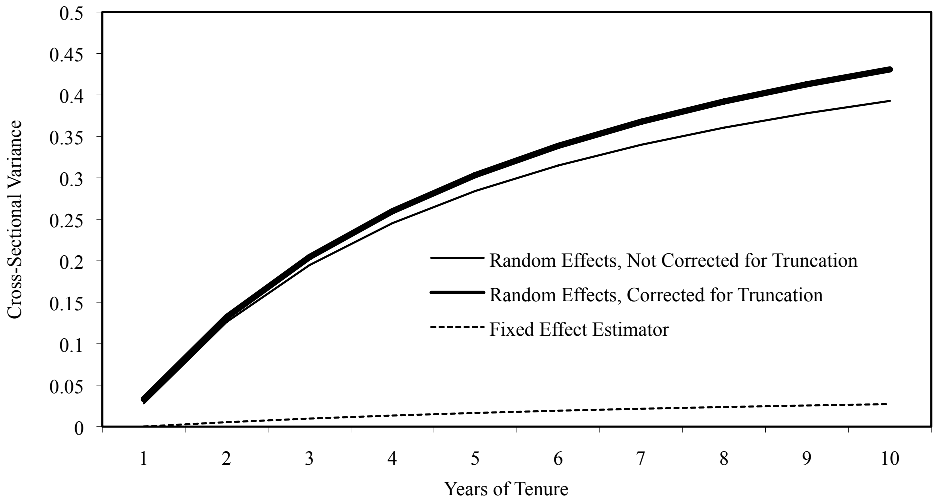
	Fixed Effect Estimator		Random Effect Estimator	
	Parameter Estimate[b]	Standard Error[c]	Parameter Estimate[b]	Standard Error[c]
No Correction for Truncation				
Variance of Measurement Error (σ^2_v)	0.202	(0.001)	0.093	(0.001)
Variance of Match Quality (σ^2_c)[a]	0.056	(0.006)	0.580	(0.018)
Variance of Initial Signal (σ^2_z)[a]	27.0	(81.2)	11.4	(1.51)
Variance of Production Outcomes (σ^2_e)[a]	0.547	(0.137)	2.55	(0.117)
p-value from Test of Overidentifying Restrictions	0.040		0.078	
Corrected for Truncation				
Variance of Measurement Error (σ^2_v)			0.122	(0.009)
Variance of Match Quality (σ^2_c)[a]			0.672	(0.107)
Variance of Initial Signal (σ^2_z)[a]			12.9	(2.83)
Variance of Production Outcomes (σ^2_e)[a]			3.48	(1.50)
p-value from Test of Overidentifying Restrictions			0.034	

[a] Estimates are based on scale parameter $\gamma=1$. To rescale the estimates for any other $0<\gamma<1$, divide the reported pararameter estimate by γ^2.

[b] Average of parameter estimates across three completed data implicates.

[c] Square root of total variance of parameter estimate across three completed data implicates, as defined in Rubin (1987).

Figure 4
Cross-Sectional Variance of the Mean of Beliefs About Match Quality

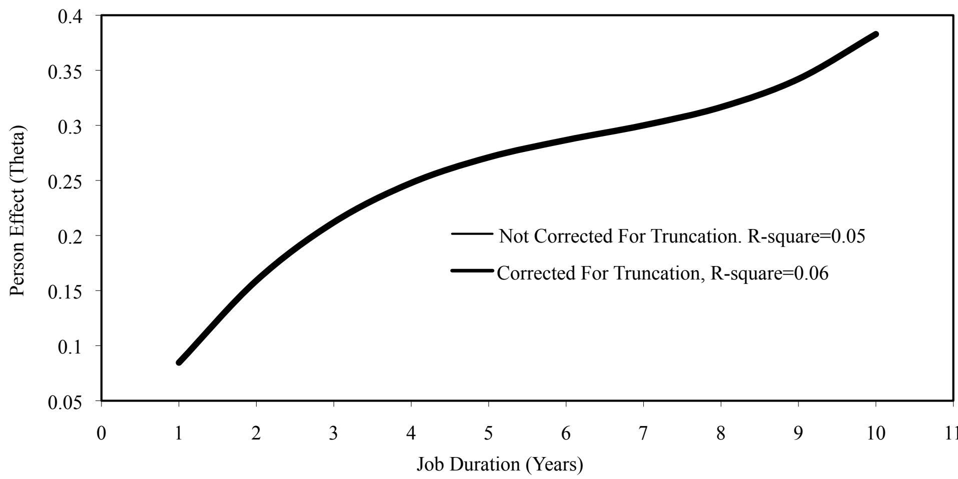


but reject after applying the truncation correction.[38]

Nevertheless, the estimates in Table 7 imply that learning about match quality is an important source of earnings dispersion. The truncation-corrected estimate of the variance of match quality is 0.672, which exceeds the estimated variance of person and firm effects in this specification. However, because agents do not directly observe match quality, this overstates its importance in earnings dispersion. That is, Bayesian learning implies that agents' point estimates of match quality (m_τ), which enter directly into wages, are less dispersed than match quality itself (except asymptotically). Consequently $V_\tau = Var\left[m_\tau\right]$ is a better measure of the importance of *learning* about match quality in earnings dispersion. This is plotted in Figure 4 for various tenure levels. Focusing on the truncation-corrected estimates, we see that V_τ increases tenfold from a modest 0.033 at one year of tenure to 0.303 by the fifth year, and 0.431 by the tenth year. Given the job tenure distribution in our data,

[38] p-Values for the test of over-identifying restrictions are based on formulae in Rubin (1987) for multiply-imputed data. Let d_m denote the test statistic from the m^{th} implicate, asymptotically distributed χ^2_k. Let M denote the number of implicates, $\bar{d}_m$ the sample mean of the statistics d_m, and s^2_d their sample variance. Define $\hat{r}_m = \left(1+M^{-1}\right)s^2_d / \left(2\bar{d}_m + \left(4\bar{d}^2_m - 2ks^2_d\right)^{1/2}\right)$ and $\hat{v} = (M-1)\left(1+\hat{r}_m^{-1}\right)^2$ .The quantity $\hat{r}_m$ is a method of moments estimator of the relative increase in variance of the test statistic due to missing data. The test statistic $\hat{D}_m = \left(\bar{d}_m/k - \frac{M-1}{M+1}\hat{r}_m\right)/\left(1+\hat{r}_m\right)$ has an asymptotic F distribution with k and $\left(1+k^{-1}\right)\hat{v}/2$ degrees of freedom. Reported p-values are based on $\hat{D}_m$.

Figure 5
Quartic Regression of Person Effect on Completed Job Duration
Random Effects With Unrestricted Within-Match Error Covariance
Left-Censored Spells Excluded (N=44,062 Completed Jobs)

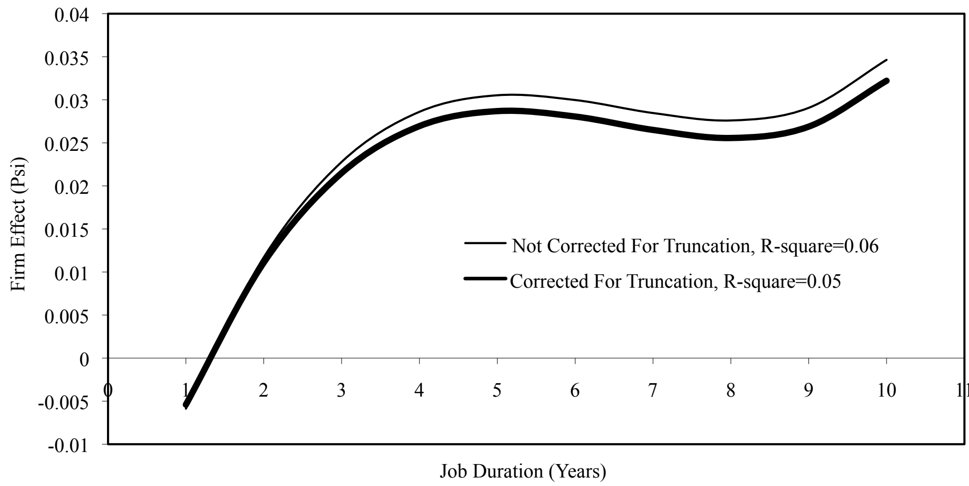


the sample average dispersion of belief terms is $\bar{V}_\tau = 0.183$. This implies a substantial portion of earnings dispersion is attributable to dispersion in beliefs about match quality. Recall that 34.9 percent of earnings variation was left unexplained by person effects, firm effects, and observables. Attributing this proportionately to measurement error ($\sigma_\nu^2 = 0.122$) and $\bar{V}_\tau$ implies that dispersion in beliefs about match quality accounts on average for 21 percent of earnings variation in our data. This exceeds the variation due to observable characteristics and firm heterogeneity, and is second only to the importance of personal heterogeneity.

5.3 Additional Predictions From the Matching Model

The matching model also predicts that larger values of θ_i and ψ_j should be associated with longer average job duration. To test this prediction I fit a fourth-order polynomial in completed job duration to the estimated person and firm effects. Figures 5 and 6 present the fitted curves for the random effect specifications with W unrestricted. Results from the other specifications are very similar and available on request. As predicted, larger values of θ_i and ψ_j are associated with longer duration. The profile for the firm effect is initially very steep, but flattens out after 4-5 years.

Figure 6
Quartic Regression of Firm Effect on Completed Job Duration
Random Effects With Unrestricted Within-Match Error Covariance
Left-Censored Spells Excluded (N=44,062 Completed Jobs)



6 Conclusion

The matching model presented in Section 2 predicts rich dispersion in equilibrium wages and employment dynamics. Productivity differences across individuals, technological differences between firms, and learning about match quality all contribute. The empirical results suggest that productivity differences across individuals are the most substantial component of earnings dispersion, accounting for 51 percent of observed variation. Learning about match quality makes a substantial contribution, with 21 percent of variation attributable to inter-match dispersion in beliefs, whereas inter-firm differences in compensation account for only 6.3 percent.

Formal and informal tests of the matching model's empirical predictions yield somewhat mixed results. On the one hand, correlations between the truncation correction term and the person and firm effects have the predicted sign, as does the relationship between the estimated person and firm effects and job duration. On the other hand, we reject the error structure implied by learning about match quality, and the estimated correlation between the person and firm effects has the wrong sign. It is worth noting, however, that other studies, particularly those on European data, have found correlations between person and firm effects in line with the model's predictions. Even in the LEHD data, close inspection of the distribution of person and firm effects reveals some evidence of mismatch as predicted by the model. Furthermore, even though we reject the learning hypothesis, the within-match error covariance appears to have, at least approximately, the martingale structure implied by Bayesian learning. It is possible that a richer learning model, e.g., where workers and firms learn one another's type slowly, might provide a better fit to the data.

Appendix A Proofs

Proof of Proposition 2. Define the operator T by

$$\begin{aligned} & (TW)(m_\tau) \\ = \; & \max\left\{ \begin{array}{c} \mu + a + b + m_\tau + \rho(1-\delta)\int W(m_{\tau+1})\, dF(m_{\tau+1}|m_\tau, s_\tau^2) \\ +\rho\delta U, U \end{array} \right\} \end{aligned}$$

where $F(m_{\tau+1}|m_\tau, s_\tau^2)$ denotes the normal distribution with mean m_τ and variance $v_{\tau+1}$ defined in equations (11) and (12), and where the second state variable s_τ^2 has been suppressed to simplify the notation. Let S denote the space of bounded, continuous, non-decreasing, convex functions. We first show $T : S \to S$. Let $q(m_\tau) = \mu + a + b + m_\tau$. The boundedness assumption implies $q(m_\tau)$ is bounded, and it is obviously continuous, increasing, and convex. Since U is a constant, it is sufficient to show that the operator M defined by

$$(MW)(m_\tau) = \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right)$$

maps S into itself. So let $W \in S$. Then MW is bounded and continuous. To see that it is non-decreasing, let $m'_\tau > m_\tau$. Then $F(m_{\tau+1}|m'_\tau, \sigma^2) \leq F(m_{\tau+1}|m_\tau, \sigma^2)$ for every $\sigma^2 > 0$. That is, $F(m_{\tau+1}|m'_\tau, \sigma^2)$ first-order stochastically dominates $F(m_{\tau+1}|m_\tau, \sigma^2)$, so that

$$\begin{aligned} (MW)(m'_\tau) &= \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m'_\tau, \sigma^2\right) \\ &\geq \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, \sigma^2\right) = (MW)(m_\tau) \end{aligned}$$

since W is non-decreasing by hypothesis. As for convexity, since $m_{\tau+1} \sim N(m_\tau, v_{\tau+1})$ we can write $m_{\tau+1} = m_\tau + \varphi$ where $\varphi \sim N(0, v_{\tau+1})$. Then rewrite the operator M as

$$(MW)(m_\tau) = \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) = \int W(m_\tau + \varphi)\, dF\left(\varphi|0, s_\tau^2\right)$$

where $F\left(\varphi|0, s_\tau^2\right)$ is the normal distribution with mean zero and variance $\upsilon_{\tau+1}$. Then for any m_τ, m'_τ, and $\lambda \in [0,1]$,

$$\begin{aligned}
& \lambda (MW)(m_\tau) + (1-\lambda)(MW)(m'_\tau) \\
= {} & \lambda \int W(m_\tau + \varphi)\, dF\left(\varphi|0, s_\tau^2\right) + (1-\lambda) \int W(m'_\tau + \varphi)\, dF\left(\varphi|0, s_\tau^2\right) \\
= {} & \int \left[\lambda W(m_\tau + \varphi) + (1-\lambda) W(m'_\tau + \varphi)\right] dF\left(\varphi|0, s_\tau^2\right) \\
\geq {} & \int W\left[\lambda (m_\tau + \varphi) + (1-\lambda)(m'_\tau + \varphi)\right] dF\left(\varphi|0, s_\tau^2\right) \\
= {} & \int W\left[\lambda m_\tau + (1-\lambda) m'_\tau + \varphi\right] dF\left(\varphi|0, s_\tau^2\right) \\
= {} & (MW)\left(\lambda m_\tau + (1-\lambda) m'_\tau\right)
\end{aligned}$$

where the inequality follows because W is convex. Thus $M : S \to S$ and hence $T : S \to S$ also.

To show there is a unique $W \in S$ that satisfies the Bellman equation (29) we need only establish that T is a contraction. Uniqueness then follows immediately from the Contraction Mapping Theorem, since S and the sup norm define a complete metric space. To show that T is a contraction, we verify the Blackwell (1965) sufficient conditions.

Monotonicity: Let $\Psi, W \in S$ and $\Psi(x) \leq W(x)$ for all x. Then

$$\begin{aligned}
(T\Psi)(m_\tau) &= \max \left\{ \begin{array}{c} q(m_\tau) + \rho(1-\delta) \int \Psi(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) \\ +\rho\delta U, U \end{array} \right\} \\
&\leq \max \left\{ \begin{array}{c} q(m_\tau) + \rho(1-\delta) \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) \\ +\rho\delta U, U \end{array} \right\} \\
&= (TW)(m_\tau)
\end{aligned}$$

as required.

Discounting: Let $W \in S$, $y \geq 0$, and $\rho \in (0,1)$. Then

$$\begin{aligned}
& [T(W+y)](m_\tau) \\
= {} & \max \left\{ \begin{array}{c} q(m_\tau) + \rho(1-\delta) \int \left[W(m_{\tau+1}) + y\right] dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) \\ +\rho\delta U, U \end{array} \right\} \\
= {} & \max \left\{ \begin{array}{c} q(m_\tau) + \rho(1-\delta) \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) \\ +\rho\delta U + \rho(1-\delta) y, U \end{array} \right\} \\
\leq {} & \max \left\{ \begin{array}{c} q(m_\tau) + \rho(1-\delta) \int W(m_{\tau+1})\, dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) \\ +\rho\delta U, U \end{array} \right\} + \rho y \\
= {} & (TW)(m_\tau) + \rho y
\end{aligned}$$

as required. ■

Proof of Proposition 3. The proof of Proposition 2 showed that when W is continuous and non-decreasing in m_τ, $\int W\left(m_{\tau+1}, s_{\tau+1}^2\right) dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right)$ is also. Therefore the first argument of the max operator in the Bellman equation (29) is continuous and increasing in m_τ. The second argument is a constant, and the result follows immediately. ■

Proof of Proposition 4. We first establish an intermediate result. Namely, that

$$\varrho\left(m_\tau, s_\tau^2\right) = E_\tau\left[W\left(m_{\tau+1}, s_{\tau+1}^2\right)\right] = \int W\left(m_{\tau+1}, s_{\tau+1}^2\right) dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right)$$

is increasing in s_τ^2. We already know from the proof of Proposition 2 that ϱ is non-decreasing in m_τ. Recall that $F\left(m_{\tau+1}|m_\tau, s_\tau^2\right)$ is the normal distribution with mean m_τ and variance $v_{\tau+1}$ given by equation (11). Notice that

$$\frac{\partial v_{\tau+1}}{\partial s_\tau^2} = \frac{s_\tau^2}{s_\tau^2 + \sigma_e^2}\left(1 + \frac{\sigma_e^2}{s_\tau^2 + \sigma_e^2}\right) > 0$$

so an increase in s_τ^2 constitutes a mean-preserving spread on $m_{\tau+1}$. Since W is convex in its first argument, for any $\tilde{s}_\tau^2 > s_\tau^2$ we have

$$\begin{aligned}
\varrho\left(m_\tau, \tilde{s}_\tau^2\right) &= \int W\left(m_{\tau+1}, s_{\tau+1}^2\right) dF\left(m_{\tau+1}|m_\tau, \tilde{s}_\tau^2\right) \\
&\geq \int W\left(m_{\tau+1}, s_{\tau+1}^2\right) dF\left(m_{\tau+1}|m_\tau, s_\tau^2\right) = \varrho\left(m_\tau, s_\tau^2\right)
\end{aligned}$$

as required.

As for the main result, suppose to the contrary that $\bar{m}_{\tau+1} < \bar{m}_\tau$. Then from (30),

$$\begin{aligned}
\bar{m}_{\tau+1} - \bar{m}_\tau &= \rho\left(1-\delta\right) \bar{E}_\tau\left[W\left(m_{\tau+1}, s_{\tau+1}^2\right)\right] \\
&\quad -\rho\left(1-\delta\right) \bar{E}_{\tau+1}\left[W\left(m_{\tau+2}, s_{\tau+1}^2\right)\right] \\
&= \rho\left(1-\delta\right)\left(\varrho\left(\bar{m}_\tau, s_\tau^2\right) - \varrho\left(\bar{m}_{\tau+1}, s_{\tau+1}^2\right)\right).
\end{aligned}$$

The right-hand side is non-negative because ϱ is non-decreasing in its first argument and $\bar{m}_{\tau+1} < \bar{m}_\tau$ by hypothesis, and because ϱ is increasing in its second argument and $s_{\tau+1}^2 < s_\tau^2$ for all $\tau > 0$. But the left-hand side is negative, a contradiction. ■

The following lemmata are useful for the proof of Proposition 5.

Lemma 6 $\partial U/\partial b = 0, \partial U/\partial a \in \left(0, \frac{1}{1-\rho}\right).$

Proof. Write the value of the worker's outside option as:

$$U = h + \rho\pi \int_{b_0}^{b_1} J_0 dF_b^* + \rho(1-\pi) U = \frac{h + \rho\pi \int_{b_0}^{b_1} J_0 dF_b^*}{1 - \rho(1-\pi)} \tag{51}$$

where

$$J_0 = E_0 [\max\{J_1, U\}] = U + \int_{\bar{m}_1}^{\infty} (J_1 - U) dF \left(m_1 | 0, \sigma_c^2 + \sigma_z^2\right) \tag{52}$$

is the prior expected value of employment defined in (18) and F_b^* is defined Appendix B. That $\partial U/\partial b = 0$ is obvious, since U doesn't depend on b.

From (51) we have

$$\frac{\partial U}{\partial a} = \frac{\rho\pi}{1-\rho(1-\pi)} \int_{b_0}^{b_1} \frac{\partial J_0}{\partial a} dF_b^*. \tag{53}$$

Differentiating (52) using Leibniz's Rule,

$$\begin{aligned} \frac{\partial J_0}{\partial a} &= \frac{\partial U}{\partial a} - \frac{\partial \bar{m}_1}{\partial a} \left(\bar{J}_1 - U\right) f\left(\bar{m}_1 | 0, \sigma_c^2 + \sigma_z^2\right) + \int_{\bar{m}_1}^{\infty} \frac{\partial (J_1 - U)}{\partial a} dF_1 \\ &= \frac{\partial U}{\partial a} + \int_{\bar{m}_1}^{\infty} \frac{\partial (J_1 - U)}{\partial a} dF_1 \end{aligned} \tag{54}$$

where $F_{\tau+1}$ is shorthand for $F\left(m_{\tau+1} | m_\tau, s_\tau^2\right)$ and $\bar{J}_\tau$ is shorthand for the value of J_τ when $m_\tau = \bar{m}_\tau$. Note $\bar{J}_\tau = U$ by definition of $\bar{m}_\tau$ and the individual rationality property of the Nash Bargain. Differentiating (16) gives the recursion

$$\frac{\partial (J_\tau - U)}{\partial a} = \gamma - \gamma(1-\rho)\frac{\partial U}{\partial a} + \rho(1-\delta) \int_{\bar{m}_{\tau+1}}^{\infty} \frac{\partial (J_{\tau+1} - U)}{\partial a} dF_{\tau+1} \tag{55}$$

for all $\tau > 0$. Substituting (55) into (54) repeatedly gives

$$\frac{\partial J_0}{\partial a} = \frac{\partial U}{\partial a} (1 - \gamma(1-\rho) Z) + \gamma Z \tag{56}$$

where

$$Z = \sum_{\tau=1}^{\infty} [\rho(1-\delta)]^{\tau-1} \int_{\bar{m}_1}^{\infty} \int_{\bar{m}_2}^{\infty} \cdots \int_{\bar{m}_\tau}^{\infty} dF_\tau \cdots dF_2 dF_1 > 0. \tag{57}$$

Substituting (56) into (53) and simplifying gives

$$\frac{\partial U}{\partial a}=\frac{1}{1-\rho}\left[\frac{\rho\pi\gamma\int_{b_0}^{b_1} ZdF_b^*}{1+\rho\pi\gamma\int_{b_0}^{b_1} ZdF_b^*}\right]\in\left(0,\frac{1}{1-\rho}\right) \tag{58}$$

because $Z>0$ implies the term in square brackets is between zero and one, and $\rho\in(0,1)$. ■

Lemma 7 *The joint value of continuing the employment relationship, $J_\tau+\Pi_\tau$, is non-decreasing in a and b.*

Proof. Since

$$J_\tau+\Pi_\tau=\mu+a+b+m_\tau+\rho(1-\delta)\int W\left(m_{\tau+1},s_{\tau+1}^2\right)dF\left(m_{\tau+1}|m_\tau,s_\tau^2\right)+\rho\delta U,$$

and given lemma 6, it is sufficient to show the value function W is non-decreasing in a and b. Recall the space of functions S from the proof of Proposition 2, and define the space of functions $S'\subseteq S$ that are also non-decreasing in a and b. Recall also the operators T and M defined in the proof of Proposition 2. Write the value function as $W(m_\tau;a,b)$. If $W\in S'$, then for any $a'>a$, $b'>b$, and any m_τ, we have $W(m_\tau;a',b')\geq W(m_\tau;a,b)$ so that

$$\begin{aligned}(MW)(m_\tau;a',b') &= \int W(m_{\tau+1};a',b')\,dF\left(m_{\tau+1}|m_\tau,s_\tau^2\right)\\ &\geq \int W(m_{\tau+1};a,b)\,dF\left(m_{\tau+1}|m_\tau,s_\tau^2\right)\\ &= (MW)(m_\tau;a,b)\end{aligned}$$

which shows that $M:S'\to S'$. Applying lemma 6 again, it follows that $(TW)(m_\tau;a',b')\geq(TW)(m_\tau;a,b)$ also. Hence $T:S'\to S'$. Since S' is a closed subset of S, the unique fixed point of T is $W\in S'$. ■

Proof of Proposition 5. Rewrite the threshold value of beliefs as

$$\bar{m}_\tau=(1-\rho)U-\mu-a-b-\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\left(J_{\tau+1}+\Pi_{\tau+1}-U\right)d\bar{F}_{\tau+1} \tag{59}$$

where $\bar{F}_{\tau+1}=F\left(m_{\tau+1}|\bar{m}_\tau,s_\tau^2\right)$. Let $x\in\{a,b\}$. Differentiating (59) using

Leibniz's Rule,

$$\begin{aligned}
\frac{\partial \bar{m}_\tau}{\partial x} &= (1-\rho)\frac{\partial U}{\partial x} - 1 \\
&\quad +\rho(1-\delta)\frac{\partial \bar{m}_{\tau+1}}{\partial x}\left(\bar{J}_{\tau+1}+\bar{\Pi}_{\tau+1}-U\right) f\left(\bar{m}_{\tau+1}|\bar{m}_\tau, s_\tau^2\right) \\
&\quad -\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\left[\frac{\partial\left(J_{\tau+1}+\Pi_{\tau+1}-U\right)}{\partial x}\right] d\bar{F}_{\tau+1} \\
&\quad -\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\frac{\partial \bar{m}_\tau}{\partial x}\left(J_{\tau+1}+\Pi_{\tau+1}-U\right)\left(\frac{m_{\tau+1}-\bar{m}_\tau}{s_\tau^2}\right) d\bar{F}_{\tau+1} \\
&= \frac{(1-\rho)\frac{\partial U}{\partial x} - 1 - \rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\frac{\partial(J_{\tau+1}+\Pi_{\tau+1}-U)}{\partial x} d\bar{F}_{\tau+1}}{1+\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\left(J_{\tau+1}+\Pi_{\tau+1}-U\right)\left(\frac{m_{\tau+1}-\bar{m}_\tau}{s_\tau^2}\right) d\bar{F}_{\tau+1}} \qquad (60)
\end{aligned}$$

where $\bar{J}_{\tau+1}$ is shorthand for the value of $J_{\tau+1}$ when $m_{\tau+1}=\bar{m}_{\tau+1}$, $\bar{\Pi}_{\tau+1}$ is defined analogously, and $\bar{J}_{\tau+1}+\bar{\Pi}_{\tau+1}=U$ by definition of $\bar{m}_{\tau+1}$.

Applying the first result from Lemma 6,

$$\frac{\partial \bar{m}_\tau}{\partial b} = \frac{-1-\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\frac{\partial(J_{\tau+1}+\Pi_{\tau+1})}{\partial b} d\bar{F}_{\tau+1}}{1+\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\left(J_{\tau+1}+\Pi_{\tau+1}-U\right)\left(\frac{m_{\tau+1}-\bar{m}_\tau}{s_\tau^2}\right) d\bar{F}_{\tau+1}}. \qquad (61)$$

Since $\partial\left(J_{\tau+1}+\Pi_{\tau+1}\right)/\partial b \geq 0$ by Lemma 7, the numerator is negative. The denominator is positive because $J_{\tau+1}+\Pi_{\tau+1} \geq U$ for $m_{\tau+1} \geq \bar{m}_{\tau+1}$ (with equality only when $m_{\tau+1}=\bar{m}_{\tau+1}$); and $m_{\tau+1} \geq \bar{m}_{\tau+1} \geq \bar{m}_\tau$ by Proposition 4. Thus $\partial \bar{m}_\tau/\partial b < 0$.

Letting $x=a$ in (60) gives

$$\frac{\partial \bar{m}_\tau}{\partial a} = \frac{(1-\rho)\frac{\partial U}{\partial a} - 1 - \rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\frac{\partial(J_{\tau+1}+\Pi_{\tau+1}-U)}{\partial a} d\bar{F}_{\tau+1}}{1+\rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty}\left(J_{\tau+1}+\Pi_{\tau+1}-U\right)\left(\frac{m_{\tau+1}-\bar{m}_\tau}{s_\tau^2}\right) d\bar{F}_{\tau+1}}. \qquad (62)$$

As in (61), the denominator is positive. To sign the numerator note that for all $s \geq 1$,

$$\begin{aligned}
J_{\tau+s}+\Pi_{\tau+s}-U &= \mu+a+b+m_{\tau+s}-(1-\rho)U \\
&\quad +\rho(1-\delta)\int_{\bar{m}_{\tau+s+1}}^{\infty}\left(J_{\tau+s+1}+\Pi_{\tau+s+1}-U\right) dF_{\tau+s+1}
\end{aligned}$$

and differentiating gives the recursion

$$\begin{aligned}
\frac{\partial\left(J_{\tau+s}+\Pi_{\tau+s}-U\right)}{\partial a} &= 1-(1-\rho)\frac{\partial U}{\partial a} \qquad (63) \\
&\quad +\rho(1-\delta)\int_{\bar{m}_{\tau+s+1}}^{\infty}\frac{\partial\left(U-J_{\tau+s+1}-\Pi_{\tau+s+1}\right)}{\partial a} dF_{\tau+s+1}.
\end{aligned}$$

Repeated substitution of (63) into the numerator of (62) gives

$$\frac{\partial \bar{m}_\tau}{\partial a} = \frac{\left[(1-\rho)\frac{\partial U}{\partial a} - 1\right] \bar{Z}_\tau}{1 + \rho(1-\delta)\int_{\bar{m}_{\tau+1}}^{\infty} (J_{\tau+1} + \Pi_{\tau+1} - U)\left(\frac{m_{\tau+1} - \bar{m}_\tau}{s_\tau^2}\right) d\bar{F}_{\tau+1}} \tag{64}$$

where

$$\bar{Z}_\tau = 1 + \sum_{s=1}^{\infty} [\rho(1-\delta)]^s \int_{\bar{m}_{\tau+1}}^{\infty} \int_{\bar{m}_{\tau+2}}^{\infty} \cdots \int_{\bar{m}_{\tau+s}}^{\infty} dF_{\tau+s} \cdots dF_{\tau+2} d\bar{F}_{\tau+1} > 0. \tag{65}$$

The numerator of (64) is negative, since $\partial U / \partial a < \frac{1}{1-\rho}$ by lemma 6. ■

Appendix B The Steady State

B.1 Flows Into Unemployment

Let $l(a, b, \tau)$ denote the density of type a workers employed at type b firms with tenure τ. The number of such workers entering unemployment in a given period is

$$\begin{aligned} & (1-u)\, l(a,b,\tau)\left[\Pr(m_\tau < \bar{m}_\tau) + \delta\right] \\ = \; & (1-u)\, l(a,b,\tau)\left[\Phi\left(\frac{\bar{m}_\tau}{V_\tau^{1/2}}\right) + \delta\right] \end{aligned} \tag{66}$$

where Φ denotes the standard normal CDF. The flow into unemployment of all type a workers from type b firms is

$$(1-u)\sum_{\tau=1}^{\infty} l(a,b,\tau)\left[\Phi\left(\frac{\bar{m}_\tau}{V_\tau^{1/2}}\right) + \delta\right] \tag{67}$$

and the aggregate flow into unemployment is

$$(1-u)\int_{a_0}^{a_1}\int_{b_0}^{b_1}\sum_{\tau=1}^{\infty} l(a,b,\tau)\left[\Phi\left(\frac{\bar{m}_\tau}{V_\tau^{1/2}}\right) + \delta\right] dbda. \tag{68}$$

B.2 Vacancies

In the steady state, the probability λ that a vacancy is filled is constant. Thus the equilibrium number of vacancies opened by each firm, i.e., the solution to (21), is also constant. Let υ_b^* denote the steady state number of vacancies opened by a type b firm, i.e.,

$$\upsilon_b^* = \arg\max_{\upsilon \in \mathbb{N}} \sum_{l=0}^{\upsilon} \binom{\upsilon}{l} \lambda^l (1-\lambda)^{\upsilon - l} \left[l \int_{a_0}^{a_1} \Pi_0 dF_a^* - \kappa(l) \right] - k_0 \upsilon \tag{69}$$

when λ takes its steady state value and where F_a^* is the steady state distribution of unemployed worker types defined below. Let f_b denote the density function associated with the distribution F_b of firm types. The steady state number of vacancies opened by all type b firms is $\phi \upsilon_b^* f_b(b)$, and the steady state stock of vacancies in the economy is

$$\upsilon = \phi \int_{b_0}^{b_1} \upsilon_b^* f_b(b) \, db. \tag{70}$$

B.3 Steady State Type Distributions

Each open vacancy is associated with a firm type b. Let $f_b^*(b)$ denote the steady state distribution of firm types among open vacancies. This is

$$f_b^*(b) = \phi \frac{\upsilon_b^*}{\upsilon} f_b(b) \tag{71}$$

with corresponding CDF F_b^*. Workers use F_b^* to compute the expected value of employment in new matches before the identity of the matching firm is known.

Similarly, we can define the distribution F_a^* of unemployed worker types. Firms use F_a^* to compute the expected value of employment in new matches before the identity of the matching worker is known. Define the density of employed type a workers:

$$l(a) = \int_{b_0}^{b_1} \sum_{\tau=1}^{\infty} l(a, b, \tau) \, db. \tag{72}$$

Then the density function f_a^* associated with F_a^* is

$$f_a^*(a) = u^{-1} \left[f_a(a) - (1-u) \, l(a) \right], \tag{73}$$

where f_a is the density function associated with the distribution F_a of worker types.

B.4 Flows Out of Unemployment

The flow of type a workers out of unemployment and into type b firms is

$$m(u,v)\, f_a^*(a)\, f_b^*(b). \tag{74}$$

Thus the aggregate flow out of unemployment is

$$m(u,v)\int_{a_0}^{a_1}\int_{b_0}^{b_1} f_a^*(a)\, f_b^*(b)\, dbda = m(u,v). \tag{75}$$

The steady state flow-balance condition is the equality of (67) and (74) for all worker types a and all firm types b. This implies the aggregate steady state flow-balance $(68) = (75)$. The steady state level of unemployment is implicitly defined by this equality when v takes its steady state value.

B.5 Firm size

Let

$$l(b) = \int_{a_0}^{a_1}\sum_{\tau=1}^{\infty} l(a,b,\tau)\, da \tag{76}$$

be the density of employment at type b firms. Then the average size of type b firms is

$$\frac{(1-u)\, l(b)}{\phi f_b(b)}. \tag{77}$$

Appendix C Data

C.1 Sample Construction

The sample is restricted to full-time private sector employees at their dominant employer,[39] between 25 and 65 years of age, who had no more than 44 employers in the sample period,[40] with real annualized earnings between \$1,000 and \$1,000,000 (1990 dollars), employed in non-agricultural jobs that included at least one full quarter of employment.[41] The sample consists of 174

[39] I identify a dominant employer for each individual in each year. An individual's dominant employer in year t is the employer at which her reported UI earnings were largest that year. About 87 percent of the UI wage records correspond to employment at a dominant employer.

[40] An extreme number of employment spells may reflect measurement error in the person and firm identifiers. Around 0.5 percent of quarterly wage observations corresponded to individuals with more than 44 employers over the sample period.

[41] An individual employed at firm j in quarter q is defined to have worked a full quarter if she was employed at j in quarters $q-1$ and $q+1$.

million quarterly earnings observations on 9.3 million individuals employed at approximately 575,000 firms, for a total of over 15 million unique worker-firm matches. The quarterly records are annualized, yielding a sample of 49.3 million annual records.

C.1.1 The Dense Samples

The dense sampling algorithm of Woodcock (2005) ensures that individuals are connected to a specified minimum number of other workers by means of a common employer. In brief, this is achieved by sampling firms first, with probabilities proportional to employment in a reference period. Workers are then sampled within firms, with probabilities inversely proportional to firm employment. A minimum of n employees are sampled from each firm. By careful choice of sampling probabilities, the dense sample is equivalent to a simple random sample of workers employed in the reference period (that is, each worker has an equal probability of being sampled), but guarantees that each worker is connected to at least n others by a common employer.

I draw two disjoint one percent dense random samples of workers employed in 1997. Each worker is connected to at least $n = 5$ others.[42] For comparison, I also draw a one percent simple random sample of workers employed in 1997. Table 1 presents connectedness properties of the full sample, Dense Sample 1, and the simple random sample.[43] The full sample is highly connected: the largest connected group contains 99.06 percent of jobs. The dense sample remains quite highly connected: about 92 percent of jobs are contained in the two largest connected groups. This is in contrast to the simple random sample: though about 80 percent of jobs are contained in the two largest groups, only 84 percent are in groups containing at least 5 worker-firm matches. By construction, all jobs in the dense samples are contained in groups of at least 5 matches. In the simple random sample, fully 5.5 percent of jobs are connected to no other.

C.2 Variable Creation and Missing Data Imputation

Missing data items include full-time status, education, tenure (for left-censored job spells), initial experience, and (in some cases discussed below) the earnings measure. Missing data items are multiply-imputed using the Sequential Regression Multivariate Imputation (SRMI) method. See Rubin (1987) for

[42] The other parameters used to draw the dense samples, defined in Woodcock (2005), are $m = 0.5$ and $p = 0.004$.

[43] Characteristics of the two dense samples are virtually identical.

a general treatment of multiple-imputation; the SRMI technique is due to Raghunathan et al. (2001); Abowd and Woodcock (2001) generalize SRMI to the case of longitudinal linked data. SRMI imputes missing data in a sequential and iterative fashion on a variable-by-variable basis. Each missing data item is multiply-imputed with draws from the posterior predictive distribution of an appropriate generalized linear model under a diffuse prior. Estimates of all imputation regressions are available on request. I generate three imputed values of each missing data item. The result is three versions of the analysis sample ("implicates"), each containing different imputed values for each missing data item.

C.2.1 Real Annualized Earnings

Real annualized earnings are constructed from real full-quarter earnings. Full quarter earnings are defined as follows. For individuals who worked a full quarter at firm j in t, full-quarter earnings equal reported UI earnings (about 80 percent of the analysis sample). For individuals who did not work a full quarter in t, one of two earnings measures is used. If the individual worked at least one full quarter in the four previous or subsequent quarters, and if reported earnings in quarter t were at least 80 percent of average real earnings in the full quarters,[44] reported earnings are treated as full-quarter earnings (12.5 percent of the analysis sample). If on the other hand reported earnings are less than 80 percent of average real average earnings in the full quarters, earnings are imputed to the full-quarter level (7.5 percent of the analysis sample). The imputation model is a linear regression on log real full quarter earnings. Conditioning variables include up to four leads and four lags of full quarter earnings (where available), year and quarter dummies, race, education (5 categories), labor market experience (linear through quartic terms), and SIC division. Separate imputation models were estimated for men and for women. For each quarter in which earnings are imputed to the full-quarter level, three imputed values are drawn from the posterior predictive distribution under a diffuse prior. Real full-quarter earnings are then annualized.

C.2.2 Education

Education is multiply-imputed from the 1990 Decennial Census long form. The imputation model is an ordered logit. There are 13 outcome categories, corre-

[44] The 80 percent cutoff was chosen as follows. For individuals that worked a full quarter in q, the median ratio or quarter q earnings to average full-quarter earnigns in quarters $q-4$ to $q+4$ was 0.8.

sponding to 0 through 20 years of education. Conditioning variables include age (10 categories), vintiles of real annual earnings at the dominant employer in 1990 or the year the individual first appeared in the sample, and SIC division. Separate imputation models were estimated for men and for women. For each person, three imputed values are drawn from the normal approximation (at the mode) to the posterior predictive distribution under a diffuse prior. The education measure is subsequently collapsed to five categories: Less than high school, High school graduate, Some college or vocational training, Undergraduate degree, and Graduate or professional degree.

C.2.3 Labor Market Experience

In the first quarter that an individual appears in the sample, I calculate potential labor market experience as the greater of age at the beginning of the quarter minus years of education minus 6, and zero. In each subsequent quarter, labor market experience is accumulated using the individual's realized labor market history.

C.2.4 Tenure

All jobs with positive earnings in the first quarter of available data for that state are presumed left-censored (33 percent of jobs). For non-left-censored spells, tenure is set to 1 in the first quarter that there is a UI wage record, and is subsequently accumulated using the individual's employment history. For left-censored spells, tenure as of the first quarter of 1990 is imputed using data from the 1996 and 1998 CPS February supplements. The imputation model is a linear regression on the natural logarithm of job tenure. Conditioning variables include age (10 categories), vintiles of real annual earnings at the dominant employer in 1990, education (5 categories), and SIC division. For each left-censored job, three imputed values of tenure in 1990 quarter 1 were drawn from the posterior predictive distribution under a diffuse prior. In subsequent quarters, tenure is accumulated using the individual's employment history.

C.2.5 Full-Time Status

Full-time status is multiply-imputed using the 1982-1999 CPS March supplements. The imputation model is a binary logit. Conditioning variables include a quadratic in age, SIC division, year dummies, and vintiles of reported annual earnings at the dominant employer. Separate imputation models were estimated for men and for women. For each worker-firm match in each year, three imputed values were drawn from the normal approximation (at the mode) to the posterior predictive distribution under a diffuse prior.

C.3 Characteristics of the Samples

Table 2 presents basic summary statistics for the full analysis sample, the two dense samples, and a simple random sample. The dense samples exhibit properties virtually identical to those of the simple random sample, confirming the analytic proof of equivalence in Woodcock (2005). Since these are point-in-time samples, their properties differ slightly from those of the full sample. In particular, they exhibit properties consistent with a sample of individuals with a somewhat stronger-than-average labor force attachment: individuals in the point-in-time samples are somewhat more likely to be male, are slightly more educated, have somewhat longer average job tenure, earn slightly more, and are somewhat more likely to work a full calendar year.

Appendix D Estimates on Log Earnings

Appendix Tables 1 through 4 replicate the estimation results from Tables 3 through 7 of the main text on log earnings. Similarly, Appendix Figures 1 through 4 replicate Figures 1 through 4 from the main text on log earnings. The estimated variance components in Appendix Table 1 are interpreted as follows. Conditional on all other effects, a one standard deviation increase in α_i increases earnings by σ_α log points. Similarly, a one standard deviation increase in ψ_j increases earnings by σ_ψ log points. In general, the estimates based on log earnings are qualitatively very similar to those reported in the main text for earnings levels.

APPENDIX TABLE 1
ESTIMATED VARIANCE COMPONENTS, REGRESSIONS ON LOG EARNINGS
Combined Results From 3 Completed Data Implicates

	Fixed Effect Estimator		Random Effects With Spherical Error		Random Effects With Unrestricted Within-Match Error Covariance	
	Estimate[a,c]	Std Error[b]	Estimate[a]	Std Error[b]	Estimate[a,d]	Std Error[b]
All Employment Spells						
No Correction for Truncation						
Variance of person effect (σ^2_α)	0.290	(0.002)	0.230	(0.005)	0.177	(0.002)
Variance of firm effect (σ^2_ψ)	0.077	(0.000)	0.153	(0.002)	0.076	(0.007)
Residual variance (σ^2_ε)	0.061	(0.000)	0.044	(0.001)	n/a	n/a
Corrected for Truncation						
Variance of person effect (σ^2_α)			0.229	(0.005)	0.176	(0.002)
Variance of firm effect (σ^2_ψ)			0.153	(0.002)	0.077	(0.006)
Residual variance (σ^2_ε)			0.044	(0.001)	n/a	n/a
Truncation Correction (β_λ)			0.042	(0.002)	0.021	(0.001)
Left-Censored Spells Excluded						
No Correction for Truncation						
Variance of person effect (σ^2_α)	0.293	(0.002)	0.236	(0.006)	0.181	(0.003)
Variance of firm effect (σ^2_ψ)	0.078	(0.000)	0.151	(0.003)	0.077	(0.001)
Residual variance (σ^2_ε)	0.065	(0.000)	0.049	(0.001)	n/a	n/a
Corrected for Truncation						
Variance of person effect (σ^2_α)			0.236	(0.005)	0.181	(0.003)
Variance of firm effect (σ^2_ψ)			0.151	(0.003)	0.076	(0.001)
Residual variance (σ^2_ε)			0.049	(0.001)	n/a	n/a
Truncation Correction (β_λ)			0.004	(0.003)	0.017	(0.003)

[a] Arithmetic mean of parameter estimate across three completed data implicates.

[b] Square root of total variance of parameter estimate over three completed data implicates, as defined in Rubin (1987).

[c] Sample variance of estimated person and firm effects, averaged over three completed data implicates.

[d] Estimates based on multiply-imputed tenure for left-censored spells.

APPENDIX TABLE 2
CORRELATIONS AMONG ESTIMATED EFFECTS, LOG EARNINGS
Combined Results From 3 Completed Data Implicates

No Correction for Truncation						
Fixed Effect Estimator	y	θ	α	Uη	ψ	Xβ
Log Earnings (y)	1					
Total Person Effect (θ)	0.74	1				
Unobserved Component (α)	0.66	0.91	1			
Observed Component (Uη)	0.34	0.41	0.00	1		
Total Firm Effect (ψ)	0.45	0.03	0.00	0.09	1	
Time-Varying Covariates (Xβ)	0.18	-0.30	-0.27	-0.12	0.05	1
Random Effects With Spherical Error						
Log Earnings (y)	1					
Total Person Effect (θ)	0.80	1				
Unobserved Component (α)	0.71	0.91	1			
Observed Component (Uη)	0.38	0.41	-0.01	1		
Total Firm Effect (ψ)	0.47	0.03	-0.01	0.08	1	
Time-Varying Covariates (Xβ)	0.29	0.02	-0.03	0.11	0.04	1
Random Effects With Unrestricted Error Covariance						
Log Earnings (y)	1					
Total Person Effect (θ)	0.82	1				
Unobserved Component (α)	0.73	0.87	1			
Observed Component (Uη)	0.36	0.49	-0.01	1		
Total Firm Effect (ψ)	0.54	0.22	0.20	0.09	1	
Time-Varying Covariates (Xβ)	0.30	0.02	-0.03	0.09	0.04	1
Corrected For Truncation						
Random Effects With Spherical Error						
Log Earnings (y)	1					
Total Person Effect (θ)	0.80	1				
Unobserved Component (α)	0.71	0.91	1			
Observed Component (Uη)	0.38	0.41	-0.01	1		
Total Firm Effect (ψ)	0.48	0.03	0.00	0.08	1	
Time-Varying Covariates (Xβ)	0.29	0.02	-0.03	0.12	0.05	1
Truncation Correction ($\beta_\lambda\lambda$)	-0.06	-0.01	-0.01	-0.02	-0.18	0.01
Random Effects With Unrestricted Error Covariance						
Log Earnings (y)	1					
Total Person Effect (θ)	0.82	1				
Unobserved Component (α)	0.73	0.87	1			
Observed Component (Uη)	0.36	0.49	-0.01	1		
Total Firm Effect (ψ)	0.54	0.22	0.20	0.09	1	
Time-Varying Covariates (Xβ)	0.30	0.02	-0.03	0.09	0.04	1
Truncation Correction ($\beta_\lambda\lambda$)	-0.06	0.00	0.00	-0.01	-0.25	0.00

APPENDIX TABLE 3
WITHIN-MATCH ERROR COVARIANCE, REGRESSIONS ON LOG EARNINGS
Left-Censored Spells Excluded, Combined Results From 3 Implicates,

Fixed Effect Estimator

Tenure	1	2	3	4	5	6	7	8	9	10
1	.088									
2	.022	.062								
3	.004	.019	.052							
4	-.004	.005	.017	.049						
5	-.010	-.002	.005	.016	.047					
6	-.015	-.008	-.002	.005	.018	.052				
7	-.015	-.010	-.004	.001	.008	.018	.047			
8	-.016	-.011	-.007	-.003	.002	.007	.019	.050		
9	-.017	-.013	-.009	-.005	-.002	.001	.010	.020	.052	
10	-.017	-.013	-.011	-.007	-.005	-.002	.005	.012	.022	.060

Random Effect Estimator

Tenure	1	2	3	4	5	6	7	8	9	10
1	.124									
2	.066	.094								
3	.052	.066	.094							
4	.047	.059	.070	.099						
5	.041	.053	.063	.075	.098					
6	.036	.049	.058	.068	.078	.104				
7	.031	.045	.054	.064	.072	.083	.107			
8	.030	.043	.052	.063	.071	.078	.086	.115		
9	.032	.043	.051	.062	.068	.075	.081	.091	.118	
10	.030	.041	.051	.061	.066	.072	.077	.083	.092	.119

Random Effect Estimator, Corrected for Truncation

Tenure	1	2	3	4	5	6	7	8	9	10
1	.160									
2	.076	.131								
3	.057	.081	.127							
4	.050	.069	.086	.134						
5	.043	.061	.074	.093	.132					
6	.038	.054	.067	.082	.097	.139				
7	.033	.049	.061	.075	.087	.103	.141			
8	.031	.046	.058	.073	.083	.094	.106	.150		
9	.032	.046	.056	.070	.079	.088	.096	.110	.153	
10	.030	.044	.055	.069	.076	.083	.090	.098	.110	.154

APPENDIX TABLE 4
ESTIMATED STRUCTURAL PARAMETERS, REGRESSIONS ON LOG EARNINGS
Combined Results From 3 Completed Data Implicates, Left-Censored Spells Excluded

	Fixed Effect Estimator		Random Effect Estimator	
	Parameter Estimate[b]	Standard Error[c]	Parameter Estimate[b]	Standard Error[c]
No Correction for Truncation				
Variance of Measurement Error (σ^2_v)	0.053	(0.000)	0.039	(0.001)
Variance of Match Quality (σ^2_c)[a]	0.010	(0.000)	0.090	(0.004)
Variance of Initial Signal (σ^2_z)[a]	0.969	(0.024)	0.097	(0.011)
Variance of Production Outcomes (σ^2_e)[a]	0.072	(0.002)	0.147	(0.052)
p-value from Test of Overidentifying Restrictions	< 0.0001		0.056	
Corrected for Truncation				
Variance of Measurement Error (σ^2_v)			0.058	(0.006)
Variance of Match Quality (σ^2_c)[a]			0.115	(0.005)
Variance of Initial Signal (σ^2_z)[a]			0.156	(0.017)
Variance of Production Outcomes (σ^2_e)[a]			0.234	(0.038)
p-value from Test of Overidentifying Restrictions			0.070	

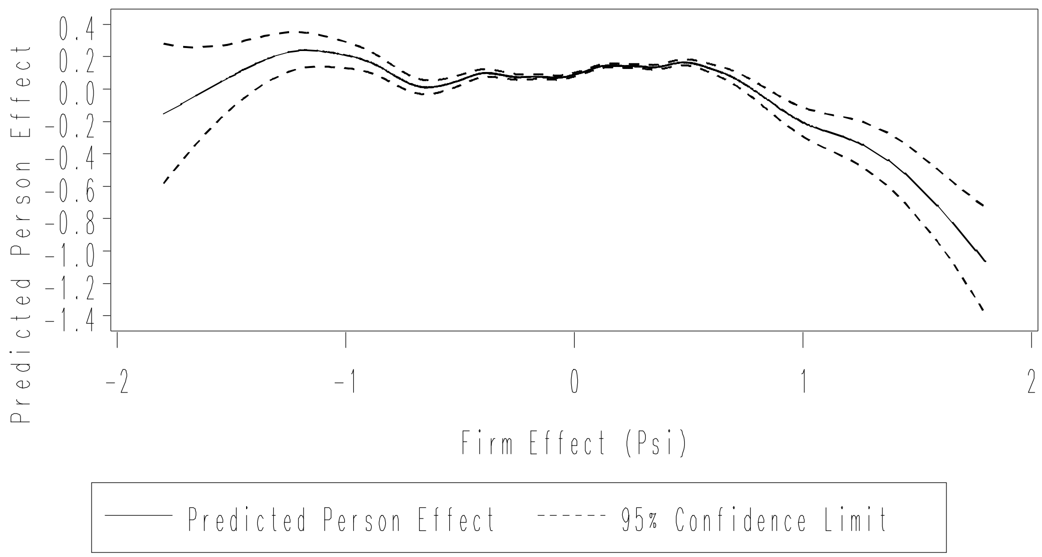
[a] Estimates are based on scale parameter γ=1. To rescale the estimates for any other 0<γ<1, divide the reported pararameter estimate by γ^2.

[b] Average of parameter estimates across three completed data implicates.

[c] Square root of total variance of parameter estimate across three completed data implicates, as defined in Rubin (1987).

Appendix Figure 1

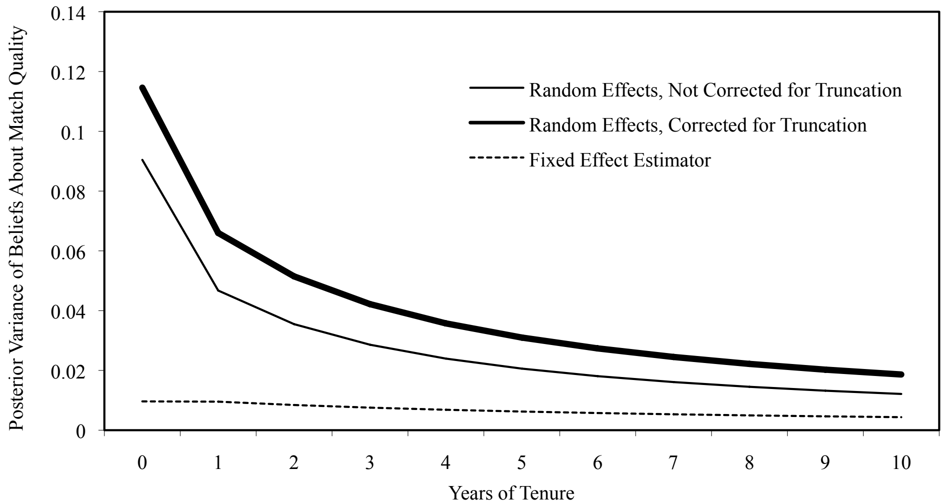
Estimated Regression of Person Effect (Theta) on Firm Effect (Psi)
Random Effects With Spherical Errors, Corrected for Truncation, Based on Log(Earnings)



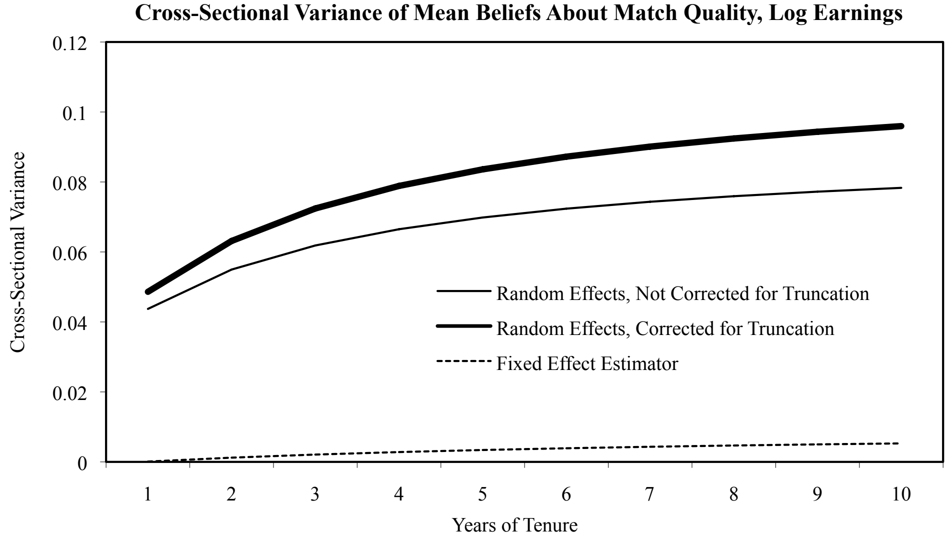
Appendix Figure 2

Estimated Regression of Person Effect (Theta) on Firm Effect (Psi)
Random Effects With W Unrestricted, Corrected for Truncation, Based on Log(Earnings)

Appendix Figure 3
Estimated Sequence of Belief Variances, Log Earnings



Appendix Figure 4
Cross-Sectional Variance of Mean Beliefs About Match Quality, Log Earnings



References

Abowd, J. M. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*(2), 411–445.

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002). Computing person and firm effects using linked longitudinal employer-employee data. Mimeo.

Abowd, J. M., F. Kramarz, P. Lengermann, and S. Perez-Duarte (2004). Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States. Mimeo.

Abowd, J. M., F. Kramarz, P. Lengermann, and S. Roux (2005). Persistent inter-industry wage differences: Rent sharing and opportunity costs. Mimeo.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–334.

Abowd, J. M., P. Lengermann, and K. McKinney (2003). The measurement of human capital in the U.S. economy. Mimeo.

Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. D. Woodcock (2009). The LEHD infrastructure files and the creation of the quarterly workforce indicators. In T. Dunne, J. B. Jensen, and M. J. Roberts (Eds.), *Producer Dynamics: New Evidence from Micro Data*, pp. 149–236. Cambridge: NBER.

Abowd, J. M. and S. D. Woodcock (2001). Disclosure limitation in longitudinal linked data. In P. Doyle, J. I. Lane, J. M. Theeuwes, and L. M. Zayatz (Eds.), *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, Chapter 10, pp. 215–277. Elsevier Science.

Abraham, K. G. and H. S. Farber (1987). Job duration, seniority, and earnings. *The American Economic Review 77*(3), 278–297.

Albrecht, J. and S. Vroman (2002). A matching model with endogenous skill requirements. *International Economic Review 43*(1), 283–305.

Alexander, R. A., K. P. Carson, G. M. Alliger, and L. Carr (1987). Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. *Educational and Psychological Measurement 47*, 309–315.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics 116*(1), 313–350.

Altonji, J. G. and R. A. Shakotko (1987). Do wages rise with job seniority? *Review of Economic Studies 54*(3), 437–459.

Andersson, F., H. J. Holzer, and J. I. Lane (2005). *Moving Up or Moving On: Who Advances in the Low-Wage Labour Market*. Russell Sage Foundation.

Andrews, M., T. Schank, and R. Upward (2004). Practical estimation methods for linked employer-employee data. IAB Discussion Paper No. 3/2004.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A 171*(3), 673–697.

Bartel, A. P. and G. J. Borjas (1981). Wage growth and job turnover: An empirical analysis. In S. Rosen (Ed.), *Studies in Labor Markets*, pp. 65–90. Chicago: National Bureau of Economic Research.

Barth, E. and H. Dale-Olsen (2003). Assortative matching in the labour market? Stylised facts about workers and plants. Institute for Social Research Working Paper.

Blackwell, D. (1965). Discounted dynamic programming. *Annals of Mathematical Statistics 36*, 226–235.

Bureau of Labor Statistics (1997). *BLS Handbook of Methods*. U.S. Department of Labor.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica 74*, 323–364.

Chamberlain, G. (1984). Panel data. In Z. Grilliches and M. Intriligator (Eds.), *Handbook of Econometrics, Volume II*, Chapter 22, pp. 1248–1318. New York: Elsevier.

Dostie, B. (2005). Job turnover and the returns to seniority. *Journal of Business and Economic Statistics 23*(2), 192–199.

Farber, H. S. and R. Gibbons (1996). Learning and wage dynamics. *The Quarterly Journal of Economics 111*(4), 1007–1047.

Felli, L. and C. Harris (1996). Learning, wage dynamics, and firm-specific human capital. *Journal of Political Economy 104*(4), 838–868.

Flinn, C. J. (1986). Wages and the job mobility of young workers. *Journal of Political Economy 94*(3), S88–S110.

Gautier, P. A. (2002). Unemployment and search externalities in a model with heterogeneous jobs and workers. *Economica 69*(273), 21–40.

Gibbons, R., L. F. Katz, T. Lemieux, and D. Parent (2005). Comparative advantage, learning, and sectoral wage discrimination. *Journal of Labor Economics 23*(4), 681–724.

Gilmour, A. R., R. Thompson, and B. R. Cullis (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics 51*, 1440–1450.

Goux, D. and E. Maurin (1999). Persistence of interindustry wage differentials: A reexamination using matched worker-firm panel data. *Journal of Labor Economics 17*(3), 492–533.

Gruetter, M. and R. Lalive (2004). The importance of firms in wage determination. IZA Discussion Paper No. 1367.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

Harris, M. and B. Holmstrom (1982). A theory of wage dynamics. *Review of Economic Studies XLIX*, 315–333.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica 46*(6), 1251–1271.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Henderson, C., O. Kempthorne, S. Searle, and C. V. Krosigk (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*(2), 192–218.

Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics 24*(1), 255–286.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy 87*(5), 972–990.

Kohns, S. (2000). Different skill levels and firing costs in a matching model with uncertainty - an extension of Mortensen and Pissarides (1994). IZA Discussion Paper No. 104.

Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics 25*, 1–35.

McCulloch, C. E. and S. R. Searle (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley and Sons.

Mincer, J. and B. Jovanovic (1981). Labor mobility and wages. In S. Rosen (Ed.), *Studies in Labor Markets*, pp. 21–64. Chicago: National Bureau of Economic Research.

Mortensen, D. (1994). The cyclical behavior of job and worker flows. *Journal of Economic Dynamics and Control 18*, 1121–1142.

Moscarini, G. (2003). Skill and luck in the theory of turnover. Yale University Working Paper.

Moscarini, G. (2005). Job matching and the wage distribution. *Econometrica 73*, 481–516.

Nagypal, E. (2007). Learning by doing vs. learning about match quality: Can we tell them apart? *Review of Economic Studies 74*, 537–566.

Newey, W. K. (1985). Generalized method of moments specification tests. *Journal of Econometrics 29*, 229–256.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(4), 2295–2350.

Raghunathan, T. E., J. M. Lepkowski, J. V. Hoewyk, and P. Solenberger (2001). A multivariate technique for multiply-imputing missing values using a sequence of regression models. *Survey Methodology 27*(1), 85–95.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science 6*(1), 15–32.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Sattinger, M. (1995). Search and the efficient assignment of workers to jobs. *International Economic Review 36*(2), 283–302.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika 78*(4), 719–727.

Searle, S. R. (1987). *Linear Models for Unbalanced Data*. New York: John Wiley and Sons.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. New York: John Wiley and Sons.

Shimer, R. (2005). The assigment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy 113*(5), 996–1025.

Shimer, R. and L. Smith (2000). Assortative matching and search. *Econometrica 68*(2), 343–369.

Shimer, R. and L. Smith (2001). Matching, search, and heterogeneity. *B.E. Journals in Macroeconomics 1*(1), Article 5.

Stern, S. (1990). The effects of firm optimizing behaviour in matching models. *Review of Economic Studies 57*, 647–660.

Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. *Quarterly Journal of Economics 107*(2), 439–479.

Woodcock, S. D. (2005). Sampling connected histories from longitudinal linked data. Mimeo.

Woodcock, S. D. (2007). Match effects. SFU Discussion Paper dp07-13.