Endogenous Mobility

John M. Abowd
Department of Economics
Cornell University
John.Abowd@cornell.edu

Ian M. Schmutte
Department of Economics
Terry College of Business
University of Georgia
schmutte@uga.edu

February 21, 2012

THIS VERSION PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE

Abstract

We establish a method of correcting for endogeneity bias in fixed-effects estimates of worker and firm-specific earnings heterogeneity using longitudinal employer-employee data. The problem arises because realized job assignments may be endogenous if they are partially determined by unobservable components of earnings. We exploit the network structure of the data to model the process by which the labor market selects job matches from the universe of possible employment relationships. Specifically, we model the evolution of the matched data as an evolving bipartite graph in a Bayesian latent class framework. We estimate the model using data from the LEHD program of the U.S. Census Bureau. Our results suggest that the correction for endogeneity is meaningful but does not overturn the qualitative findings from previous analyses that assumed mobility to be exogenous.

This paper was written while the first author was Distinguished Senior Research Fellow and the second author was RDC Administrator at the U.S. Census Bureau. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer-Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Abowd also acknowledges direct support from NSF Grants SES-0339191, CNS-0627680, SES-0922005, TC-1012593, and SES-1131848.

1 Introduction

The objective of this paper is to explore the consequences for estimates of worker and firm effects in the decomposition of earnings in longitudinally-linked employer-employee data when there is endogenous mobility, and to propose methods of correcting those estimates. Abowd et al. (1999) pioneered the identification, computation, and inference for the fixed effect estimator of the decomposition of log earnings into components associated with unobserved worker and employer heterogeneity. A major factor in the interpretation of their statistical model is that it requires the assumption that the assignment of workers to firms is random conditional on all observable characteristics and the design of the stationary unobservables. This assumption is at odds with many models of job assignment, in particular, those in which workers sort into jobs according to their comparative advantage. Since structural interpretations of the measured heterogeneity have major consequences for our understanding of the labor market, it is important to weaken the assumption that job mobility and assignments are exogenous to earnings.

The central problem of this paper is one manifestation of a fundamental challenge of empirical social science: separating the influence of correlated unobservables and sorting from the direct effect of group membership. Our approach is analogous to estimating treatment effects in the presence of selection on unobservables. Here the complication is that the number of possible treatments, that is employers, is in the millions. We construct an instrument for the actual assignment of workers to firms that exploits the relational structure of our data. The key insight is that the work histories of one's coworkers and previous employers are informative of one's own employment history, while being plausibly unrelated to whatever idiosyncratic wage innovations drive assignment at the margin.

Correcting estimated firm effects on wages for endogeneity bias is useful in a number of applications. First, this paper contributes to the ongoing debate as to whether estimates of employer-specific wage premia constitute evidence in contradiction of the law of one wage. If the bias correction does not affect the overall significance of firm effects in the presence of worker effects, it suggests that firms really do play an important direct role in wage determination, consistent with sociological evidence, but contrary to the neoclassical model of a competitive labor market. Furthermore, it helps to resolve some of the debates spawned by the early empirical results based on the assumption of exogenous mobility. These early results show little correlation between estimated employer wage premia and worker-specific earnings. In other words, high wage workers do not systematically appear in high-wage firms. This has often been cited as evidence against theories of assortative matching, and in favor of models of frictional search, which predict this lack of assortativity. The empirical results have spawned a theoretical literature attempting to construct frictional search models with assortative matching, in which estimated employer and worker wage components would misrepresent the true assignment structure in the economy (Abowd, Kramarz, Pérez-Duarte and Schmutte 2009; Shimer 2005; Lentz 2010).

Estimates of individual and firm effects are also being increasingly used in downstream applications. Iranzo et al. (2008) and Abowd et al. (2003) use estimates of person effects to measure the human capital distribution within firms. Combes et al. (2008) use a similar decomposition to estimate the contribution of neighborhoods to spatial earnings dispersion. Schmutte (2010) relies on consistent estimates of firm effects to infer the role of local job referral networks on earnings outcomes. Our estimates should be of interest in all such applications, as the specter of endogenous mobility clouds the interpretation of empirical results that rely on consistency of estimated individual or firm effects.

There is also a parallel literature in the economics of education that uses value-added models to estimate the contributions of teachers and classrooms to student achievement. Endogenous assignment is just as much a problem for those models as it is here. Indeed, several recent studies have shown that the assumption of exogenous assignment in value-added models is rejected by the data (Rothstein 2010; Koedel and Betts 2010). Nevertheless, validation studies have shown that estimates from the value-added models are significantly correlated with independent assessments of teacher productivity. Our techniques provide a direct method of assessing whether correcting the endogeneity bias in value-added models would substantially change their results. Our method can be implemented as long as one has data in which the realized network of connections between individuals and groups is sufficiently detailed to provide identifying variation.

We proceed by setting up the log earnings decomposition proposed by Abowd et al. (1999) so that we can clearly articulate the nature of the endogenous mobility problem. Then, we report the results of two tests of the exogenous mobility assumption conducted by Abowd et al. (2010). As we will see, the null of exogenous mobility is rejected, but the associated analysis of mobility patterns provides interesting information about the nature of the true model. Next, we present an illustrative theoretical model with endogenous job mobility and suggest an IV estimator based on the use of relational information derived from the network structure. Turning to implementation, we set up a formal statistical model with endogenous mobility, in which earnings and job mobility are determined by latent classifications of workers, firms, and matches. This is a mixture model in which the probability of forming a link between a given worker and firm depends on a latent classification. We show that the model is identified and show how to estimate it using the Gibbs sampler.

2 Why Does Endogenous Mobility Matter?

2.1 Exogenous Mobility in the Abowd, Kramarz and Margolis Model

Abowd et al. (1999) originally proposed the linear decomposition of log wage rates as the least squares fit of the equation

$$w = X\beta + D\theta + F\psi + \varepsilon \tag{1}$$

where w is the $[N \times 1]$ stacked vector of log wage outcomes w_{it}, X is the $[N \times k]$ design matrix of observable individual and employer time-varying characteristics (the intercept is normally suppressed, with y and X measured as deviations from overall means); D is the $[N \times I]$ design matrix for the individual effects; F is the $[w \times J - 1]$ design matrix for the employer effects (non-employment is suppressed here). ε is the $[N \times 1]$ vector of statistical errors whose properties will be elaborated below; $\left[\begin{array}{ccc} \beta^T & \theta^T & \psi^T \end{array}\right]^T$ are the unknown effects with dimension $[k \times 1], [I \times 1]$, and $[J - 1 \times 1]$ associated with each of the design matrices.[1]

The assumption of exogenous mobility appears in the assumption that

$$\mathrm{E}(\varepsilon|X, D, F) = 0.$$

As long as the matrix of data moments has full rank – a non-trivial assumption – this conditional moment restriction yields a consistent estimator for the full parameter vector, including the individual and employer effects. Exogenous mobility imposes that a worker's employment history is completely independent of the idiosyncratic part of earnings captured in ε, which in the AKM model includes the "match effect" – the average amount by which log wages in the current $(i, \mathrm{J}(i,t))$ match deviate from their expected value. The AKM model is, thus, equivalent to assuming that all assignments are pre-determined at birth given full knowledge of X, D, F and $\left[\begin{array}{ccc} \beta^T & \theta^T & \psi^T \end{array}\right]^T$. Hence, there is no room for features included in many models of job mobility and assignment to affect either the duration of matches or the assignment of workers to particular employers.

Identification of

$$\left[\begin{array}{ccc} \beta^T & \theta^T & \psi^T \end{array}\right]$$

in the statistical model also requires that $[X\,D\,F]^T\,[X\,D\,F]$ be of full rank. Abowd et al. (2002) showed that this condition is equivalent to connectedness of the realized mobility network constructed by connecting sub-graphs of all workers who share a common employer and all employers who share a common worker over the entire longitudinal sample. The realized mobility network is a static bipartite graph on worker and employer nodes. As we will see, our identification strategy also has an interpretation in terms of the realized mobility network. We use information in the realized mobility network that predicts employer assignments but that we assume is conditionally independent of earnings residuals to achieve identification. The identification conditions for the least squares solution for the parameters in Equation (1) are orthogonality of each component of the design matrix with respect to the estimated residual vector, implying that the estimated effects are also orthogonal to the estimated residuals.

$$\hat{\beta}^T X^T \hat{\varepsilon} = 0,\ \hat{\theta}^T D^T \hat{\varepsilon} = 0 \text{ and } \hat{\psi}^T F^T \hat{\varepsilon} = 0 \tag{2}$$

[1] The AKM formulation is an analysis of covariance with two high-dimension factors (individuals and employers) whose levels are estimated by least squares.

2.2 Tests of Exogenous Mobility

Abowd et al. (2010) develop two formal tests for the null hypothesis of exogenous job mobility against 2 omnibus alternatives that encompass many forms of endogenous mobility. They apply these tests to longitudinally integrated employer-employee data from the Longitudinal Employer-Household Dynamics (LEHD) Program of the U.S. Census Bureau. Here we survey the basic nature of the tests and their results. Their tests exploit the implicit restriction that future assignments of workers to firms are uninformative about current earnings residuals. Under the null hypothesis of exogenous mobility, these future assignments have no predictive power with respect to the residual. The first test, “Test 1,” checks whether a worker’s future employers are independent of the average residual in the current job. The second test, “Test 2,” checks whether the future employees of a particular employer are predictive of the residuals on their current period wage payments.

Both tests reject the null of exogenous mobility. The test statistic for Test 1 is $X^2_{8,991} = 7,438,692$ with $\Pr\{X^2_{\nu_2}\} < 0.001$. The test statistic for Test 2 is $X^2_{900} = 172,295$ with $\Pr\{X^2_{\nu_2}\} < 0.001$. These are consistent with the related tests in Rothstein (2010) that reject the exogenous mobility assumption in value-added models for longitudinally linked education data.

3 A General Model of Wage Dynamics with Endogenous Mobility

Here, we set up a somewhat simplified version of that strategy. We assume workers, firms and matches have latent classifications that we cannot observe, but that determine earnings and mobility. The statistical model is very general and is compatible with many different structural models. Formally, we use the latent class model to identify workers and firms with similar mobility and earnings patterns and then to estimate the effect on earnings related to membership in these classes. We conduct Bayesian inference using an adaptation of the Gibbs sampler algorithm for finite mixture models (Tanner 1996; Diebolt and Robert 1994) to our case allowing for multiple overlapping levels of correlation across observations. Our application and proposed procedure are related to stochastic blockmodels and other methods for the detection of “communities” of nodes in social networks. Our main innovation is the use of both node and edge characteristics in predicting the matches (Hoff et al. 2002; Newman and Leicht 2007; Neville and Jensen 2005).

3.1 Model Setup

Agents of the model are workers, indexed $i \in \{1 \dots I\} \equiv I$ and firms, indexed $j \in \{0 \dots J\} \equiv J$, where $j = 0$ is “not employed.”Each worker has a latent ability class denoted $a_i \in A$ and each firm, except $j = 0$, has a latent productivity class denoted $b_j \in B$. In addition, each worker-firm match has an associated heterogeneity component that affects both wages and mobility: $k_{ij} \in K$. A, B and K are discrete with cardinality $L, M+1$ and Q. The "not-employed firm"is a single entity in its heterogeneity class, so the class b_0 has no employer heterogeneity. To make the subsequent formulas easier to interpret, assume that the elements of A, B and K are rows from the identity matrices I_L, I_M and I_Q, respectively For instance, if $L = 2$, we have $A = \{(1,0),(0,1)\}$. The assignments of workers and firms to ability and productivity classes are independent multinomial random variables with parameters π_a, π_b. We allow for endogeneity in the match quality by letting the probability of k depend on ability and productivity. So $\Pr(k_{ij} = k | a_i = a, b_j = b) = \pi_{k|ab}$.

The log of earnings, when actually employed in any match, is given by

$$w_{ijt} = \alpha + a_i\theta + b_j\psi + k_{ij}\mu + \varepsilon_{it} \tag{3}$$

where θ, ψ, μ are vectors of parameters describing the effect on the level of log earnings associated with membership in the various heterogeneity classes. We take ε to be normal with mean 0 and variance σ^2, independent and identically distributed across individuals and over time. When not employed, the individual earns a reservation log wage of

$$w_{i0t} = \alpha + a_i\theta + \psi_0 + k_{i0}\mu + \varepsilon_{it} \tag{4}$$

where ψ_0 is just the appropriate element of ψ and $k_{i0}\mu$ allows for heterogeneity in home production with the same effects as in the market sector.

We formalize endogenous mobility by allowing those matches and employment durations that are observed to depend on ability, productivity and match quality. Let $\mathrm{J}(i,t)$ be the index function that returns the identifier of the firm in which i is employed in period t. Define the variable $s_{it} = 1$ if i separates from his current job at the end of period t and $s_{it} = 0$ otherwise. We let the probability of separation depend on the match quality by specifying

$$\Pr\left(s_{it} = 1 | k_{i\,\mathrm{J}(i,t)}\right) = \mathrm{f}_{se}\left(a_i, b_{\mathrm{J}(i,t)}, k_{i\,\mathrm{J}(i,t)}; \gamma\right) \equiv \gamma_{abk} \tag{5}$$

where $0 \leq \gamma_{abk} \leq 1$. Conditional on separation, the productivity class of the next employer depends on the productivity of the current employer, the ability of the worker, and the quality of the current match

$$\Pr\left(b_{\mathrm{J}(i,t+1)} | a_i, b_{\mathrm{J}(i,t)}, k_{i\,\mathrm{J}(i,t)}\right) = \mathrm{f}_{tr}\left(a_i, b_{\mathrm{J}(i,t)}, k_{i\,\mathrm{J}(i,t)}; \delta\right) \equiv \delta_{abk} \in \Delta^{M+1} \tag{6}$$

where $\delta_{abk} \equiv \left[\delta_{0|abk}, ..., \delta_{M|abk}\right]$ is a $1 \times (M+1)$ vector of transition probabilities, Δ^{M+1} is the unit simplex, and $\mathrm{J}(i,0) = 0$ for all i. The transition probabilities are indexed by all of the latent heterogeneity in the model. Within a heterogeneity class, the identity of the precise employer selected is completely random, as is the identity of an individual within an ability class.

4 The Network Interpretation of Endogenous Mobility Models

Let the set of identifiers for all I individuals who work in one of the J employers (including non-employment), $A(t)$, and the set of J employers, $E(t)$, be arranged in a bipartite graph where $A(t)$ and $E(t)$ are the two (disjoint) vertex (or node) sets. There is a link between $i \in A(t)$ and $j \in E(t)$ if and only if i is employed by j at date t. The totality of these links active at date t can be represented by the $I \times J$ adjacency matrix $B(t)$.[2] Assuming that we are modeling primary job holders only, this adjacency matrix has a special form that is critical in the modeling.

The labor market bipartite graph summarized by $B(t)$ evolves over time. Since the employment relations between firms and workers can change at any time, it is reasonable to think of t as a continuous variable, sampled at intervals reflected in the data. These considerations motivate adopting the dynamic network modeling tools to address the endogenous mobility issues.

We distinguish primary employment from other forms of employment. The primary employer at time t is the current employer if there is only one. Otherwise, the primary employer is the one to whom the individual supplies the most labor market time. This assumption puts constraints on the row degree distribution of $B(t)$.

Specifically, assume that $j = 0$ refers to the non-employment state. Including the column $j = 0$ ensures that every individual in the population at date t has exactly one "employer" although the (shadow) log wage outcome will be unobserved for individuals who are not employed at t. Hence, $B(t)\, e_J = e_J$, where e_J is the $J + 1 \times 1$ column vector of 1s. The column degree distribution, $e_I' B(t)$, is the size distribution of employers (technically only the columns 1 to J are included in this distribution). The employer size distribution (including non-employed) is therefore the column degree distribution of the labor market bipartite graph. The (very hard) problem of entry and exit of individuals and employers can be included in this formalism by including columns in B for potential and defunct employers and allowing for birth and death of individuals. For the moment, we are not going to address this extension.

The observed labor market data are snapshots of the market at points in time,

$$B(t_1), ..., B(t_T),$$

where T is the total number of available time periods. These adjacency matrices describe outcomes sampled at discrete points in time from the $I \times J + 1$ potential outcomes at each moment of time. The objective is to use these snapshots of the labor market to model how the labor market evolves over time and to implement tests of various endogenous mobility models.

[2] Formally, the adjacency matrix for the bipartite graph of all edges in the node set $\{A(t), E(t)\}$ is block diagonal with matrices of zeros on the diagonals, the matrix $B(t)$ in the upper right block and $B(t)^T$ in the lower left block. Only $B(t)$ contains any information, so we refer to it as the adjacency matrix throughout.

The adjacency matrix representation can be directly related to the AKM framework. When the sort order of the data is t primary and i secondary, we have

$$F=\left[\begin{array}{c} B(1) \\ B(2) \\ \vdots \\ B(T) \end{array}\right]$$

where $B(t)$ is exactly the adjacency matrix from the bipartite labor market graph (with column J removed if the rows corresponding to non-employment are removed) and F is the design matrix for the employer effects in the AKM specification. A direct strategy for modeling endogenous mobility is, therefore, to model the evolution of $B(t)$. We adopt this approach below.

The evolution of $B(t)$ can be described by a continuous time stochastic process as in Snijders (2001). Such a formulation allows the dissolution of an edge (employment separation) or creation of an edge (employment accession) to occur at any time in the interval between t_s and t_{s+1}. We adopt the simpler approach of allowing at most one transition during period t that occurs at the beginning of $t+1$. Under these simplifications the evolution of $B(t)$ can be described by the Markov transition matrix $\Pr\left[b_{ij}(t+1)=1 | B(t), \{\text{Data}\}\right]$, where $\{\text{Data}\}$ is the observed and latent data from $s=0,\ldots,t$.

5 Likelihood, Prior and Posterior Distributions

5.1 Likelihood functions

We begin by developing the likelihood function for the observed and latent data. The observed data, y_{it}, consist of wage rates, separations, accessions, and identifier information:

$$y_{it}=\left[w_{i\,\mathrm{J}(i,t)t}, s_{it}, i, \mathrm{J}(i,t)\right] \text{ for } i=1,...,I \text{ and } t=1,...T. \tag{7}$$

The latent data vector, Z, consists of the heterogeneity classifications:

$$Z=\left[a_1,\ldots,a_I,b_1,\ldots,b_J,k_{11},k_{12},\ldots,k_{1J},k_{21},\ldots,k_{IJ}\right]. \tag{8}$$

In practice, we only use or update the heterogeneity classifications for the matches that are actually observed, the number of which is bounded above by $T\times I$. That is, we only care about k_{ij} where i,j is such that $j=J(i,t)$ for some t. Finally, the complete parameter vector is

$$\rho^T=\left[\alpha,\theta^T,\psi^T,\mu^T,\sigma,\gamma,\delta,\pi_a,\pi_b,\pi_{k|ab}\right], \rho\in\Theta \tag{9}$$

We assume that workers and firms are infinitely-lived. The complete process starts at $t=1$, with continuous sampling continuing to date T. We model initial conditions by assuming that everyone enters the labor force at $t=1$ and is assigned an employer completely at random. In other words, we assume that the matches initially observed are exogenous. The observed data matrix for this time interval is denoted Y. The likelihood function for the joint distribution (Y, Z) is given by

$$
\begin{aligned}
\pounds\left(\rho|Y,Z\right) \propto & \prod_{i=1}^{I}\left\{\begin{array}{c}
\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\left(w_{i\,\mathrm{J}(i,t)t}-\alpha-a_i\theta-b_{\mathrm{J}(i,t)}\psi-k_{i\,\mathrm{J}(i,t)}\mu\right)^2}{2\sigma^2}\right]\\
\times\prod_{t=1}^{T-1}\left[1-\gamma_{\langle a_i\rangle\left\langle b_{\mathrm{J}(i,t)}\right\rangle\left\langle k_{i\,\mathrm{J}(i,t)}\right\rangle}\right]^{1-s_{it}}\left[\gamma_{\langle a_i\rangle\left\langle b_{\mathrm{J}(i,t)}\right\rangle\left\langle k_{i\,\mathrm{J}(i,t)}\right\rangle}\right]^{s_{it}}\\
\times\prod_{t=1}^{T-1}\left[\delta_{\left\langle b_{\mathrm{J}(i,t+1)}\right\rangle|\langle a_i\rangle\left\langle b_{\mathrm{J}(i,t)}\right\rangle\left\langle k_{i\,\mathrm{J}(i,t)}\right\rangle}\right]^{s_{it}}
\end{array}\right\}\\
& \times\prod_{i=1}^{I}\prod_{j=1}^{J}\left[\left(\prod_{\ell=1}^{L}\prod_{m=1}^{M}\prod_{q=1}^{Q}\left(\pi_{a\ell}\right)^{a_{i\ell}}\left(\pi_{bm}\right)^{b_{jm}}\left(\pi_{q|\ell m}\right)^{k_{ijq}}\right)\right] \qquad (10)
\end{aligned}
$$

where the notation $\pi_{a\ell}$ denotes the ℓ^{th} element of π_a (similarly for π_{bm}, b_{jm}, *etc.*) and $\langle x\rangle$ means the index of the non-zero element of the vector x.

The likelihood factors into a part due to the observed data conditioned on the latent data, and the latent data conditioned on the parameters. The observed data likelihood conditional on the latent data factors further into separate contributions from the earnings and the mobility processes. The mobility process is Markov, and conditionally independent of the earnings realizations once we know the latent classifications of the workers, firms and matches. The power of the model comes from the predictive equation for Z given the observed data and the parameters, which we can compute as the complete data likelihood divided by the observed data likelihood. The observed data likelihood is calculated by integrating out the latent data.

5.2 Prior distributions

The parameter vector ρ has a prior distribution that is composed of the product of priors on each of the main components of the parameter space. Conditional on the heterogeneity probabilities, the coefficients in the log wage equation have prior distributions proportional to a constant (each one uniform on a wide, but finite, interval of $\mathbb{R}$) and subject to the constraint that the probability-weighted average effects are all zero. That is,

$$\pi_a^T\theta = \pi_b^T\psi = \pi_{k|ab(\ell m)}^T\mu = 0 \qquad (11)$$

for all ℓ, m where $\pi_{k|ab(\ell m)} \equiv \Pr\left(k_{ij} = k|a_i = \ell, b_j = m\right)$. The variance parameter, σ, has the inverted gamma prior $\mathrm{IG}\left(\nu_0, s_0\right)$ with prior degrees of freedom small and prior s_0^2 large. Each vector of probabilities has a Dirichlet prior with each element of the parameter vector given by the inverse of the dimension of the probability vector.

5.3 Posterior distributions

The posterior distribution of ρ given (Y, Z) is given by

$$p(\rho|Y,Z) \propto \pounds(\rho|Y,Z)\frac{1}{\sigma^{\nu_0+1}}\exp\left(-\frac{s_0^2}{\sigma^2}\right)\prod_{\ell=1}^{L}\pi_{a\ell}^{\frac{1}{L}-1}\prod_{m=1}^{M}\pi_{bm}^{\frac{1}{M}-1} \tag{12}$$
$$\times\prod_{\ell=1}^{L}\prod_{m=0}^{M}\prod_{q=1}^{Q}\left(\pi_{q|\ell m}^{\frac{1}{Q}-1}\gamma_{\ell mq}^{\frac{1}{2}-1}\left(1-\gamma_{\ell mq}\right)^{\frac{1}{2}-1}\prod_{m'=0}^{M}\delta_{m'|\ell mq}^{\frac{1}{M+1}-1}\right),$$

which factors into independent posterior distributions as follows:

$$\begin{bmatrix}\alpha\\ \theta\\ \psi\\ \mu\end{bmatrix}|\sigma \sim N\left(\begin{bmatrix}\hat{\alpha}\\ \theta\\ \psi\\ \mu\end{bmatrix},\sigma^2\left(G^TG\right)^{-1}\right) \tag{13}$$

where

$$\begin{bmatrix}\hat{\alpha}\\ \theta\\ \psi\\ \mu\end{bmatrix} = \left(G^TG\right)^{-1}G^Tw$$

$$\sigma^2 \sim \mathrm{IG}\left(\frac{\nu}{2},\frac{2}{\nu s^2}\right) \tag{14}$$

$$\pi_a \sim \mathrm{D}\left(n_{a1}+\left(\frac{1}{L}-1\right),\ldots,n_{aL}+\left(\frac{1}{L}-1\right)\right) \tag{15}$$

$$\pi_b \sim \mathrm{D}\left(n_{b1}+\left(\frac{1}{M}-1\right),\ldots,n_{bM}+\left(\frac{1}{M}-1\right)\right) \tag{16}$$

$$\pi_{k|ab} \sim \mathrm{D}\left(n_{k|ab1}+\left(\frac{1}{Q}-1\right),\ldots,n_{k|abQ}+\left(\frac{1}{Q}-1\right)\right) \tag{17}$$

$$\gamma_{lmq} \sim \mathrm{D}\left(n_{\ell mq}^{sep}+\left(\frac{1}{2}-1\right),n_{\ell mq}^{stay}+\left(\frac{1}{2}-1\right)\right) \tag{18}$$

$$\delta_{b|lmq} \sim \mathrm{D}\left(n_{0|\ell mq}^{trans}+\left(\frac{1}{M+1}-1\right),\ldots,n_{M|\ell mq}^{trans}+\left(\frac{1}{M+1}-1\right)\right) \tag{19}$$

These factorizations contain some additional notation. $G = [A\,B\,K]$ is the full design matrix of ability, productivity, and match types in the complete data, and w is the vector of observed earnings. The term ν in the posterior of σ is $\nu = N + \nu_0 - (L + M + Q)$ and

$$s^2 = \frac{\left(w - G\begin{bmatrix}\hat{\alpha}\\ \theta\\ \psi\\ \mu\end{bmatrix}\right)^T \left(w - G\begin{bmatrix}\hat{\alpha}\\ \theta\\ \psi\\ \mu\end{bmatrix}\right)}{\nu}. \tag{20}$$

The remaining parameters are sampled from Dirichlet posteriors, denoted by D.

Finally, we have various counts from the completed data. $n_{a\ell}$ is the count of workers with ability class ℓ. n_{bm} is the number of employers in productivity class m. $n_{k|abq}$ is the number of matches observed in quality class q. n_{lmq}^{sep} is the number of observations in which a worker in ability class ℓ separates from an employer in productivity class m when match quality was q. Finally, $n_{m'|\ell mq}^{trans}$is the number of transitions by workers in ability class ℓ from a match with an employer in productivity class m and match quality class q to an employer in productivity class m'.

6 Estimation Procedure

We start with initial values for the parameter vector and latent data, $\rho^{(0)}, Z^{(0)}$. We defined above the distributions of the parameters given the observed and latent data. To complete the specification, we define the distributions for the latent variables conditional on the observed data and the parameters. For instance, to update the ability classifications for the workers, we need to sample from a multinomial with probability of the ℓ^{th} class equal to

$$\begin{aligned} p\left(a_i = \ell | a_{-i}, b, k, Y, \rho\right) &= \frac{p\left(a_{-i}, b, k, Y | \rho, a_i = \ell\right) p(a_i = \ell)}{p\left(a_{-i}, b, k, Y | \rho\right)} \\ &= \frac{\pi_{a\ell} p\left(a_{-i}, b, k, Y | \rho, a_i = \ell\right) p(a_i = \ell)}{\sum_{\ell'=1}^{L}\left[\pi_{a\ell'} p\left(a_{-i}, b, k, Y | \rho, a_i = \ell'\right) p(a_i = \ell')\right]}. \end{aligned} \tag{21}$$

This requires computing the likelihood function under each assignment of i to an ability classification. The update formulas for b_j and k_{ij} are exactly analogous. This is a high dimension procedure, requiring roughly L evaluations of the likelihood per individual, M per firm, and Q per match, for each iteration. However, given the simple form of the likelihood, these computations are not be excessively burdensome. Furthermore, much of this work can be parallelized. The updating for each a_i is an independent task. Furthermore, most of the structure of the likelihood function remains the same as we tweak individual assignments, which we exploit to obtain further simplification.

With the posterior distributions as defined in the previous section, the Gibbs sampler can be implemented as follows:

$$\sigma^{(1)} \sim p\left(\sigma | \alpha^{(0)}, \theta^{(0)T}, \psi^{(0)T}, \mu^{(0)T}, \gamma^{(0)}, \delta^{(0)}, \pi_a^{(0)}, \pi_b^{(0)}, \pi_{k|ab}^{(0)}, Z^{(0)}, Y\right) \tag{22}$$

$$\begin{bmatrix} \alpha \\ \theta \\ \psi \\ \mu \end{bmatrix}^{(1)} \sim p\left(\begin{bmatrix} \alpha \\ \theta \\ \psi \\ \mu \end{bmatrix} | \gamma^{(0)}, \delta^{(0)}, \pi_a^{(0)}, \pi_b^{(0)}, \pi_{k|ab}^{(0)}, Z^{(0)}, \sigma^{(1)}, Y \right) \tag{23}$$

$$\gamma^{(1)} \sim p\left(\gamma | \delta^{(0)}, \pi_a^{(0)}, \pi_b^{(0)}, \pi_{k|ab}^{(0)}, Z^{(0)}, \alpha^{(1)}, \theta^{(1)T}, \psi^{(1)T}, \mu^{(1)T}, Y\right) \tag{24}$$

$$\vdots \tag{25}$$

$$k_{IJ}^{(1)} \sim p\left(k_{IJ} | \rho^{(1)}, a_1^{(1)}, \ldots, a_I^{(1)}, b_1^{(1)}, \ldots, b_J^{(1)}, k_{11}^{(1)}, \ldots, k_{IJ-1}^{(1)}, Y\right) \tag{26}$$

7 Simulation Study

We demonstrate the validity of our estimation procedure with a simulation study. Our model is an extension of standard data augmentation (Tanner 1996). As such, theoretical results on the convergence of the data augmentation algorithm and Gibbs sampler should obtain here. The main novelty in our approach is the presence of multiple levels of latent variables. Each observation belongs to three separate latent classes: a worker-specific ability class, an employer-specific productivity class, and a match-specific quality class. As we show, our Gibbs sampler performs well for inference regarding the parameters associated with earnings effects of the latent classifications.

We simulate data under a model with $L = M = Q = 2$ heterogeneity classes. In our model economy, there are $I = 100$ workers and $J = 20$ employers. We observe workers in each of $T = 50$ time periods. The simulated data allow for both the separation decision and the job allocation at transition to depend on match quality. Furthermore, match quality is correlated with latent worker ability and latent employer productivity, so there is a rich structure of endogenous mobility in the model. We simulate the model under the parameterization described in Section C.1.

Section C.2 presents summary statistics for the data simulated under the model. The data contain 373 distinct employer-employee matches, not including spells of unemployment. The simulated job mobility and earnings histories reflect the model parameterization. As expected given the model, the average of log earnings per period, calculated across periods of employment, is indistinguishable from zero.

Table 1 reports the average correlation between heterogeneity components of earnings estimated using the AKM decomposition, estimated using our structural model, across each of the 1000 draws from the Gibbs sampler. Of primary interest for our purpose are estimates of the Abowd et al. (1999) decomposition in these data. The true correlation between estimated person- and firm-effects in realized matches in these estimates is $-.063$. The correlation estimated using the AKM decomposition is $-.069$.

In this, the sampler converges within 100 iterations to point mass on the true latent classifications, regardless of choice of initial conditions. Thereafter, the sampler converges immediately to sampling from the posterior distribution for the parameters. This is because the posterior for the parameters given the latent data factors into the product of Dirichlet and Normal-Gamma posteriors. The results presented here use 1000 samples from the posterior distribution following a 1000 sample burn-in. We initialize all parameters in the wage equation to zero, and start the categorical variables at uniform distributions. The sampler iterates between sampling from the posterior distribution for the model parameters conditional on the observed and latent data, and sampling from the predictive distribution for the latent classes given the observed data and parameters. The choice of starting point does not matter: up to a relabeling of the heterogeneity classes, the sampler quickly converges to the same solution.

To sample from the posterior for $\alpha, \theta, \psi, \mu, \sigma$ requires an identifying assumption. Here, unlike Abowd et al. (2002), we set one effect equal to zero. Since we only have two classes for each type of heterogeneity, this means we estimate four parameters for the earnings equation along with the standard deviation of the structural error. In addition, we have to specify prior parameters for the Normal-Gamma distribution. We assume prior degrees of freedom equal to one, and also set the prior standard error equal to one. The prior for $(\alpha, \theta, \psi, \mu)$ is normal with mean zero and prior covariance matrix equal to the identity matrix.

Figure 1 plots of the posterior distribution of the parameters from the wage equation. By comparing these values with the truth presented in section C.2, we see that up to re-identification, the model converges to the truth. As a result, our procedure is able to detect the small endogeneity bias in the estimated correlation between worker and employer heterogeneity classes. Finally, section C presents the posterior mode for the categorical variables in the model. A comparison with the model parameters and the simulated separation, transition, and match rates show that the model is very accurately capturing the evidence from the data. Again, this is unsurprising since the simulator quickly finds the true latent ability, productivity, and match classifications for the data.

8 Structural Estimation

We implement the model empirically using matched employer-employee data from the LEHD program of the U.S. Census Bureau. The basic structure of these data is described in Abowd, Stephens, Vilhuber, Andersson, McKinney, Roemer and Woodcock (2009). The data processing and estimation procedures used to create the AKM decomposition for these data are described in Abowd et al. (2003).

8.1 Universe Used for Structural Estimation

We used a universe sample from the states of Illinois, Indiana, and Wisconsin. All individuals who worked in these states during the years 1999-2003 for their dominant job (the one with the most annual earnings) were included. There are 16.9 million persons from 719 thousand unique employers. There are 39 million unique matches. The summary statistics for these data provide the AKM starting values for our structural estimation. They are reported in Appendix D. In particular, the values of the AKM parameters θ and ψ were discretized into two categories based on the medians for the universe of persons and employers (considered as separate populations) in the three-state sample. The AKM residual was averaged over all periods in which a match was present to obtain the continuous version of the AKM μ, which was then discretized into two categories using the median of the match population. This is the procedure that was also used for the exogenous mobility tests in Abowd et al. (2010).

8.2 Sampling procedure

We fit the likelihood function in equation (10) using a 0.25% simple random sample of individuals, keeping all employers and match-years associated with those individuals. The 0.25% simple random sample of persons has 42,228 persons, 39,458 firms, 97,455 matches (including non-employment) and 211,140 person-years (also including years spent in non-employment). Because of the structure of the frame for this sample, which included only three geographically contiguous states, non-employment includes employment spells outside of the universe, and should be interpreted in that light when examining the structural estimates.

8.3 Estimation Details

The same Gibbs sampler that was tested in the Monte Carlo simulations, equation (22) through equation (26), was used to fit the structural model to the sample LEHD data. The starting values are shown in Appendix D. The results reported here use the last 500 draws from 1,000. All threads were in steady-state after 200 draws, as was the case for our Monte Carlo simulations.

8.4 Results

Figure 2 shows the posterior distribution of the structural wage equation parameters. All of these posterior distributions are tight around the modal value. Table 2 presents these results in traditional format so that it is easy to see that the posterior means are all several orders of magnitude greater than the posterior standard deviations. None of the wage function parameters associated with the structural version of the AKM decomposition has a symmetric distribution around zero.

Table 3 provides a very interesting comparison of the wage decomposition parameters estimated by least squares (labeled AKM) and our endogenous mobility model (labeled Gibbs). In computing this table, we used the continous values from the AKM decomposition but, of necessity, only the two discrete values from our structural model. The structural person effect, θ_{Gibbs}, explains less of the variance of the dependent variable (labeled y) than its AKM counterpart, while the structural firm effect, ψ_{Gibbs}, explains about the same amount. Somewhat surprisingly, the structural match component, μ_{Gibbs}, explains very little of the variation in the dependent variable and is negatively correlated with it. In the AKM estimates, the person and firm effect have a weak postive correlation of 0.0635 and in the structural estimation the correlation is also weak and positive, 0.0359. The AKM match effect, because it is estimated from the least squares residual must be essentially uncorrelated with the person and firm effects, as is the case in our subsample. However, the structural person and firm effects are weakly (person) and strongly (firm) negatively correlated with the match effect. Nothing is correlated with the residual.

Table 4 displays the regression of the structural estimates of the wage decomposition components on the AKM estimates of all components. These regressions, therefore, compute the conditional expectation of the structure given the AKM estimates. They can be use to compute endogenous mobility-corrected estimates of the wage components from data for which only the AKM estimates are available. Of course, more specification testing should be performed to confirm that the other samples share the same mobility parameters as the one we used to estimate this table.

Figure 3 presents the contrast between the population distributions of workers, firms, and matches and compares them to the steady-state distributions across realized matches. The horizontal axis presents the distribution of workers across ability types, π_A: we estimate 71 percent of workers are of low ability. The vertical axis reflects the population distribution of firms across productivity types, π_B: 53 percent of firms are of low productivity. The cells indicated by dashed black lines are therefore the populations of potential matches. Within each cell, the black number provides the probability that the match is of high quality conditional on the firm and worker type, $\pi_{K|AB}$. High productivity firms are much less likely to generate a good match, and in general, there seems to be a negative correlation between person type, firm type, and match quality.

Overlaid on the figure in blue are the steady-state distributions across realized matches. The model generates a closed-form for the transition of any worker across labor market states. We present the average of this Markov transition matrix across 500 draws from the Gibbs sampler in Table 5. We compute the kernel of this transition matrix to obtain the expected distribution of workers across firm types, match types, and non-employment.

We find that among employed workers, 68 percent are of low ability. For the low ability workers, 49 percent of their jobs are in low productivity firms. For high ability workers, 51 percent are. Among realized matches, the difference between the population and steady state distributions are slightly more dramatic. Among realized jobs, those involving low productivity employers are more likely to be of high quality than the population distribution would suggest.

9 Conclusion

Could we randomly assign workers to jobs without changing the manner in which their wages were determined? Either their employers know of the random assignment, and would, presumably, compensate them differently than workers they hired, or they would not, in which case those workers would be non-randomly selected from the pool of potential applicants. It is easy to imagine randomizing applications but not realized assignments. We do not have an ideal experiment that identifies the effect of assignments of workers to firms. It is difficult to think of what an ideal experiment would be.

The central problem of this paper is one manifestation of a fundamental challenge of empirical social science: separating the influence of correlated unobservables and sorting from the direct effect of group membership. Exploiting the wealth of information about labor market behavior locked in the relational structure of matched data holds great potential to address these problems. We use matched data to construct a complete model for the actual assignment of workers to firms that exploits the dynamic network structure of our data.

Bibliography

Abowd, J. M., Creecy, R. H. and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data, *Technical Report TP-2002-06*, LEHD, U.S. Census Bureau.

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High wage workers and high wage firms, *Econometrica* **67**(2): 251–333.

Abowd, J. M., Kramarz, F., Pérez-Duarte, S. and Schmutte, I. M. (2009). A formal test of assortative matching in the labor market. NBER Working Paper No. 15546.

Abowd, J. M., Lengermann, P. and McKinney, K. L. (2003). The measurement of human capital in the U.S. economy, *Technical Report TP-2002-09*, LEHD, U.S. Census Bureau.

Abowd, J. M., McKinney, K. and Schmutte, I. M. (2010). How important is endogenous mobility for measuring employer and employee heterogeneity? mimeo.

Abowd, J. M., Stephens, B. E., Vilhuber, L., Andersson, F., McKinney, K. L., Roemer, M. and Woodcock, S. (2009). *Producer Dynamics: New Evidence from Micro Data*, Chicago: University of Chicago Press for the National Bureau of Economic Research, chapter The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators, pp. 149–230.

Combes, P.-P., Duranton, G. and Gobillon, L. (2008). Spatial wage disparities: Sorting matters!, *Journal of Urban Economics* **63**: 723–742.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling, *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(2): 363–375.

Hoff, P. D., Raftery, A. and Handcock, M. (2002). Latent space approaches to social network analysis, *Journal of the American Statistical Association* **97**: 1090–1098.

Iranzo, S., Schivardi, F. and Tosetti, E. (2008). Skill dispersion and firm productivity: An analysis with matched employer-employee data, *Journal of Labor Economics* **26**: 247–285.

Koedel, C. and Betts, J. R. (2010). Does student sorting invalidate value-added models of teacher effectiveness? an extended analysis of the Rothstein critique, *Education Finance and Policy* **5**: 54–81.

Lentz, R. (2010). Sorting by search intensity, *Journal of Economic Theory* **145**(4): 1436–1452.

Neville, J. and Jensen, D. (2005). Leveraging relational autocorrelation with latent group models, *MRDM '05: Proceedings of the 4th international workshop on Multi-relational mining*, ACM, New York, NY, USA, pp. 49–55.

Newman, M. E. J. and Leicht, E. A. (2007). Mixture models and exploratory analysis in networks, *Proceedings of the National Academy of Sciences of the United States* **104**: 9564–9569.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay and student achievement, *Quarterly Journal of Economics* **125**(1): 175–214.

Schmutte, I. (2010). Job referral networks and the determination of earnings in local labor markets. Cornell University Department of Economics.

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

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics, *Sociological Methodology* **30**: 361–395.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer Series in Statistics, 3rd edn, Springer-Verlag.

Figures and Tables

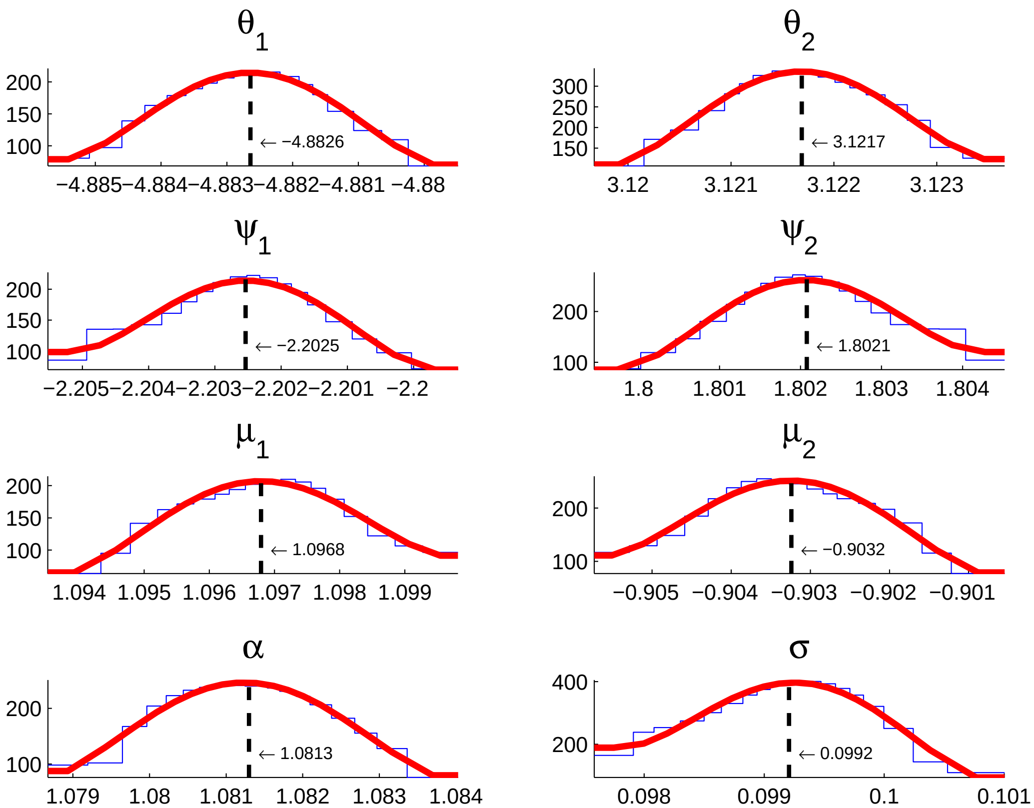

Figure 1: Posterior Distribution of Wage Equation Parameters: Simulated Data

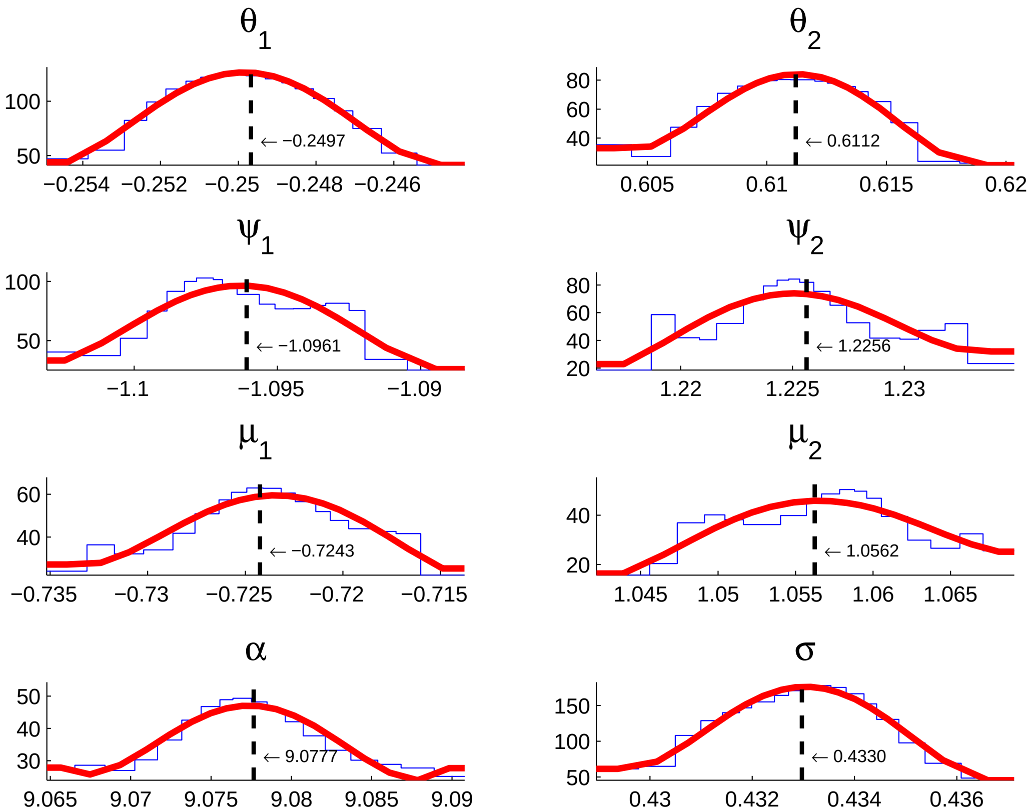

Figure 2: Posterior Distribution of Wage Equation Parameters: LEHD Data

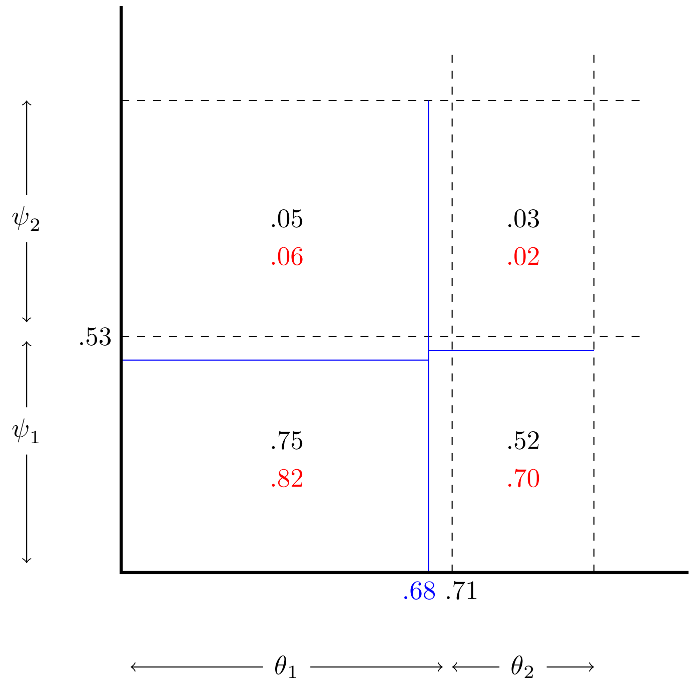

Figure 3: Population and Sampling Distributions

Table 1: Correlation Matrix of Wage Equation Parameters: Simulated Data

	y	θ_{AKM}	ψ_{AKM}	μ_{AKM}	ε_{AKM}	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}	θ_{True}	ψ_{True}	μ_{True}	ε_{True}
y	1												
θ_{AKM}	0.879	1											
ψ_{AKM}	0.393	-.069	1										
μ_{AKM}	0.141	-.000	-.000	1									
ε_{AKM}	0.024	0	0	0	1								
θ_{Gibbs}	0.867	0.985	-.066	0.000	0	1							
ψ_{Gibbs}	0.395	-.062	0.990	-.000	0	-.065	1						
μ_{Gibbs}	-.097	-.292	0.183	0.607	0	-.404	0.166	1					
ε_{Gibbs}	0.025	0.001	-.002	0.017	0.960	-.000	-.000	0.001	1				
θ_{True}	0.867	0.985	-.066	0.000	0	1	-.065	-.404	-.000	1			
ψ_{True}	0.395	-.062	0.990	-.000	0	-.065	1	0.166	-.000	-.065	1		
μ_{True}	-.097	-.292	0.183	0.607	0	-.404	0.166	1	0.001	-.404	0.166	1	
ε_{True}	0.051	0.020	0.016	0.017	0.960	0.020	0.018	-.005	0.999	0.020	0.018	-.005	1

Table entries are means of the correlation between the indicated variables across 1000 draws from the Gibbs sampler described in the text.

Table 2: Posterior Distribution of Wage Equation Parameters: LEHD Data

Parameter	Mean	Std. Dev
θ_1	-0.2497	0.0032
θ_2	0.6112	0.0051
ψ_1	-1.0961	0.0044
ψ_2	1.2256	0.0055
μ_1	-0.7243	0.0066
μ_2	1.0562	0.0082
α	9.0777	0.0082
σ	0.4330	0.0024

Table entries are means and standard deviations estimated from 500 draws from the Gibbs sampler described in the text.

Table 3: Correlation Matrix of Wage Equation Parameters: LEHD Data

	y	θ_{AKM}	ψ_{AKM}	μ_{AKM}	ε_{AKM}	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}
y	1								
θ_{AKM}	0.5284	1							
ψ_{AKM}	0.5683	0.0632	1						
μ_{AKM}	0.4236	0.0335	-.0182	1					
ε_{AKM}	0.2345	-.0000	-.0000	0.0000	1				
θ_{Gibbs}	0.3361	0.2401	0.1682	0.0816	-.0000	1			
ψ_{Gibbs}	0.5486	0.2037	0.5599	0.1179	-.0000	0.0359	1		
μ_{Gibbs}	-.02219	0.0951	-.2577	0.1396	0.0000	-.1202	-.7236 1		
ε_{Gibbs}	0.4989	0.2288	0.1498	0.2677	0.4703	-.0000	.0002	-.0000	1

Table entries are means of the correlation between the indicated variables across 500 draws from the Gibbs sampler described in the text.

Table 4: Regression of Structural Wage Decomposition Components on AKM Estimates of Wage Decomposition Components: LEHD Data

	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}
θ_{AKM}	0.1505	0.3171	0.1540	0.1747
	(.0016)	(.0039)	(.0034)	(.0017)
ψ_{AKM}	0.1442	1.4923	-.5288	0.1642
	(.0022)	(.0054)	(.0048)	(.0023)
μ_{AKM}	0.0716	0.3323	0.2656	0.3071
	(.0022)	(.0055)	(.0048)	(.0023)
ε_{AKM}	0.0000	0.0000	0.0000	0.9881
	(.0040)	(.0098)	(.0087)	(.0042)
Constant	0.0144	0.1852	-.0907	-.0034
	(.0010)	(.0024)	(.0021)	(.0010)

Results from running a regression of the wage components estimated under the endogenous mobility model on wage components estimated using the AKM decomposition. The reported values are the mean parameter estimate and standard error across 500 draws from the Gibbs sampler.

Table 5: Markov Transition Matrix: LEHD Data

θ Type							
	ψ Type		1	1	2	2	3
		μ Type	1	2	1	2	-
1	1	1	0.4127	0.2272	0.0708	0.0038	0.2855
1	1	2	0.0530	0.7230	0.0906	0.0048	0.1286
1	2	1	0.0169	0.0526	0.8447	0.0066	0.0792
1	2	2	0.0055	0.0173	0.0670	0.7530	0.1572
1	3	-	0.0552	0.1715	0.0818	0.0043	0.6872
2	1	1	0.7109	0.1014	0.0118	0.0004	0.1755
2	1	2	0.0448	0.8693	0.0405	0.0013	0.0441
2	2	1	0.0103	0.0113	0.9283	0.0038	0.0463
2	2	2	0.0123	0.0135	0.1436	0.7672	0.0634
2	3	-	0.1095	0.1202	0.0854	0.0027	0.6822

Table entries are functions of the means of the parameter estimates from 500 draws of the Gibbs sampler described in the text.

Appendices

A Endogenous Mobility in a Model with Comparative Advantage and Search Frictions

A.1 Workers and Firms

The labor market consists of two discrete and finite populations: a population of I workers, and a population of J firms. I and J are both very large. Firms and workers match to produce output for an external market. Time is discrete.

Workers differ in the skills they bring to the market, and firms differ in their ability to extract output from a given quantity of efficient labor. As well, when a worker is matched to a specific firm, the match can vary in quality in a manner that also affects productivity and output. More specifically,

- Any worker is endowed with ability, θ, drawn from the distribution $S_\theta(\theta)$ with support $[\theta_{\min}, \theta_{\max}]$;
- Any firm is endowed with productivity ψ, drawn from the distribution $S_\psi(\psi)$ with support $[\psi_{\min}, \psi_{\max}]$;
- Any worker-firm match has a quality, μ, drawn from the conditional distribution $H(\mu|\theta, \psi)$ with support $[\mu_{\min}, \mu_{\max}]$.
- $s_\theta(\theta), s_\psi(\psi), h(\mu|\theta, \psi)$ are probability densities corresponding to the distributions defined above.

Workers, firms, and matches can be nascent, active, or inactive. Workers and firms all begin in the nascent state, and must become active for at least one period before becoming inactive. Inactivity is an absorbing state.

- ν is the probability that an active worker becomes inactive at the end of the current period;
- η is the probability that an active firm becomes inactive at the end of the current period;
- δ is the probability that an active match becomes inactive at the end of the current period.

For now, assume that ν and η are also the probabilities that nascent workers and firms become active in any given period. When they are born, their types are drawn from the same distribution, and since I and J are very large, the population distribution over types for active and inactive workers and firms should be stable over time argument for this].

Log output in period t, $y_t = \ln Y_t$ in a firm-worker match is defined as

$$y_t = \psi_{J(i,t)} + \theta_i + \mu_{iJ(i,t)} \tag{27}$$

where i is the name of the worker, whose ability is θ_i. $J(i,t)$ identifies i's employer at t, the productivity of which is $\psi_{J(i,t)}$. Finally, the quality of the match between the two is $\mu_{iJ(i,t)}$.

If a match becomes inactive, but the worker in the match does not, the worker is unemployed at the start of the next period. Following Cahuc, Postel-Vinay and Robin (2006), an unemployed worked of type α receives log earnings in unemployment given by

$$\ln w_{i0} = b + \theta_i \tag{28}$$

which is equivalent to being employed by a firm with productivity b in a match with quality $\mu = 0$ and receiving all of the output as a wage.

A.2 Matching

Workers and firms come together pairwise through sequential, random, and costly search. In each period, an unemployed worker is matched with a firm for an offer with probability λ_0 and an employed worker is matched with a firm for an offer with probability λ_1. Given that a worker has an offer, the firm from which it originates is randomly selected from the population of active firms as follows:

The probability that a firm of type ψ is sampled is

$$\hat{f}_t(\psi) = \frac{f(\psi)}{\sum_{z:z=\psi_j \text{ for some j active at t}} f(z)} \tag{29}$$

for all ψ such that some active firm j is type ψ_j at time t. $\hat{f}_t(\psi) = 0$ for all ψ that are not represented among active firms at time t. f represents the search distribution by type if the full population of types are represented. This cumbersome notation is necessary to deal with the fact that we have to handle two-sided matching in a model with discrete and finite quantities of workers and firms.

A.3 Wages

Firms offer workers a constant piece rate share of the match surplus. The match surplus depends on the value of unemployment, which is constant across workers of the same ability type, and the output from the current job. The piece rate α is constant across employer types and match qualities. The log wage offered to a worker with ability θ from a firm with type ψ into a match with quality μ is therefore

$$w_{it} = \alpha + \theta_i + \psi_{J(i,t)} + \mu_{iJ(i,t)}. \tag{30}$$

This is a simplification of the equilibrium wage-posting models that this paper builds on in which firms match outside offers leading to renegotiation of the wage within a match. By focusing on a constant piece rate, we illuminate the selection issues driving endogenous mobility. We do, however, retain the implication that a worker only changes jobs to move into a better match.

A.4 Timing

At the end of the period we have the following transition probabilities

- worker (enters or) leaves the market for good with probability ν.
  - Employer (enters or) leaves the market with probability η.
  - When a worker enters, his type is drawn from $S_\theta(\theta)$.
  - When a firm enters, its type is drawn from $S_\psi(\psi)$.
  - matches exogenously dissolve with probability δ.
  - employed worker receives outside offer with probability λ_1.
  - Assume $\nu+\eta+\delta+\lambda_1 \leq 1$ (Prob. nothing happens to a worker is $[1-(\nu+\eta+\delta+\lambda_1)]$).
  - Unemployed workers receives offers with probability λ_0.

A.5 Job Mobility

Now describe the process governing the transition dynamics of workers across firms and matches. The selection of job changes has the features of the Roy model. Workers move when and only when they get an offer from job offering a higher wage. As in LeMaire and Scheuer, let the joint firm-match productivity be $\kappa = \psi + \mu$. Define the sampling distribution of firm-match quality offers to be $G(\kappa|\theta)$, which can be derived from F (search distribution of employer productivities) and H (the sampling distribution of match quality conditional on θ and ψ), which are defined above. Let $\kappa_{it} = \psi_{J(i,t)} + \mu_{iJ(i,t)}$ be the overall quality of worker i's job in period t. Let the outside job offer be given by $\kappa^* = \zeta^* + \mu^*$.

For an employed worker, given θ_i, transition dynamics over κ_{it} are simply

$$\kappa_{it+1} = \begin{cases} b & \text{with prob. } \eta+\delta; \\ \cdot & \text{with prob. } \nu; \\ \kappa_{it} & \text{with prob. } 1-\eta-\delta-\lambda_1[1-G(\kappa_{it}|\theta_i)]; \\ \kappa & \text{with density } \lambda_1 g(\kappa|\theta_i). \end{cases} \tag{31}$$

We can rewrite these in terms of firm and match heterogeneity so that the transition dynamics for an employed worker are

$$(\psi_{it+1}, \mu_{it+1}) = \begin{cases} (b,0) & \text{with prob. } \eta+\delta; \\ (\cdot,\cdot) & \text{with prob. } \nu \\ (\psi_{it}, \mu_{it}) & \text{with prob. } 1-\eta-\delta-\nu-\lambda_1[1-G(\psi_{it}+\mu_{it}|\theta_i)]; \\ (\psi,\mu) & \text{with density } \lambda_1 h(\mu|\psi,\theta_i) f(\psi). \end{cases} \tag{32}$$

For a non-employed worker, given θ_i, transition dynamics are

$$\kappa_{it+1} = \begin{cases} b & \text{with prob. } 1-\lambda_0; \\ \kappa & \text{with density } \lambda_0 g(\kappa|\theta_i). \end{cases} \tag{33}$$

In terms of firm and match heterogeneity

$$(\psi_{it+1}, \mu_{it+1}) = \begin{cases} (b,0) & \text{with prob. } 1-\lambda_0; \\ (\psi,\mu) & \text{with density } \lambda_0 h(\mu|\psi,\theta_i) f(\psi). \end{cases} \tag{34}$$

We now have enough information to completely specify wage and mobility dynamics. The model generates endogenous mobility because the correlation between θ, ψ and μ induces differential mobility. Again, for two workers of the same ability in the same firm, one with a better match will stay longer and exit to a better match. Therefore, assignment to a new employer is not independent either of the match on the 'inside' job, or of the 'outside' job.

The likelihood function for data generated under this model is presented in Appendix B.

Relative to our more general model, presented below, this model imposes a restriction on the data generating process. Namely, the model assumes that the way firms are sampled does not depend on worker quality or match quality, so all of the observed association between worker type, firm type, residual, and future employer type are generated by the Roy-style selection process, which is embedded here into the probability of job change: $\Pr(c=1|\theta,\psi,\mu)$.

A.6 Moments of interest

Define the variable $c^J_{it} = 1$ if worker i changes jobs at the end of period t. $c^J_{it} = 0$ otherwise. For employed workers, (and suppressing subscripts)

$$\Pr(c^J = 1|\theta,\psi,\mu) = \lambda_1 \left[1 - G(\psi+\mu|\theta)\right]. \tag{35}$$

For non-employed (but active) workers

$$\Pr(c^J = 1|\theta) = \lambda_0. \tag{36}$$

The sampling distribution for aggregate firm-match quality given θ is simply

$$\begin{aligned} g(k|\theta) &= \Pr(\psi, \mu|\psi + \mu = k, \theta) \\ &= \int_{-\infty}^{\infty} h(k - \psi|\theta, \psi) f(\psi) d\psi. \end{aligned} \tag{37}$$

Therefore the probability of changing jobs for a worker of type α conditional on receiving an offer is

$$\begin{aligned} 1 - G(\bar{\psi} + \bar{\mu}|\theta) &= \int_{\bar{\psi}+\bar{\mu}}^{\infty} g(k|\theta) dk \\ &= \int_{\bar{\psi}+\bar{\mu}}^{\infty} \left[\int h(k - \psi|\theta, \psi) f(\psi) d\psi \right] dk \\ &= \int \left[\int_{\bar{\psi}+\bar{\mu}}^{\infty} h(k - \psi|\theta, \psi) dk \right] f(\psi) d\psi \\ &= \int \left[1 - H\left(\bar{\psi} + \bar{\mu} - \psi|\theta, \psi \right) \right] f(\psi) d\psi \\ &= 1 - \mathbb{E}_F \left[H\left(\bar{\psi} + \bar{\mu} - \psi|\theta, \psi \right) \right] \end{aligned} \tag{38}$$

i.e. find the probability of drawing match quality sufficient to move conditional on employer type, and integrate that across the employer type distribution.

B Likelihood Function for the Correlated Matches Model

For any worker, i, the model generates the following vector of observed data:

$$y_{it} = \left[w_{it}, c_{it}, u_{it,} i, J(i,t), N(i,t) \right], \tag{39}$$

where $J(i,t)$ indicates the employer that i began the period with, $N(i,t)$ indicates the employer that i ends the period with, and c_{it} identifies whether any change of employment status took place. Specifically, $c_{it} = \max\left(c_{it}^m, c_{it}^d, c_{it}^f, c_{it}^J \right)$, where each entry can either be equal to zero or exactly one of the following:

- $c_{it}^m = 1$ if and only if i exits to non-employment because of match dissolution;
- $c_{it}^d = 1$ if and only if i becomes non-active (exits the labor market permanently);
- $c_{it}^f = 1$ if and only if i exits to non-employment because his employer becomes non-active;
- $c_{it}^J = 1$ if and only if i makes a direct job-to-job transition.

m_{it} is a binary indicator with $m_{it} = 1$ if i is employed at an observed wage during period t and 0 otherwise.

Our model also generates latent data that are observed by the market participants, but not the econometrician. The latent data are the heteogeneity classifications of each worker, firm, and match that appears in the data. The latent data vector is therefore

$$Z = [\theta_1, \ldots \theta_I, \psi_1, \ldots, \psi_J, \mu_{11}, \mu_{12}, \ldots, \mu_{1J}, \mu_{21}, \ldots, \mu_{IJ}]. \tag{40}$$

In addition to the heterogeneity already described, we assume earnings are afflicted by some kind of classical measurement error with $\varepsilon \sim N(0, \sigma^2)$, so that

$$w_{it} = \alpha + \theta_i + \psi_{J(i,t)} + \mu_{iJ(i,t)} + \varepsilon_{it}. \tag{41}$$

The parameters to be estimated are

$$\rho = [\alpha, \theta, \psi, \mu, \sigma, \lambda_0, \lambda_1, \eta, \delta, \nu, h, f, s_\psi, s_\theta]$$

The likelihood of the joint distribution of the observed and latent data is

$$\pounds\left(\rho|Y,Z\right) \propto \prod_{i=1}^{I} \left\{ \begin{array}{c} \prod_{t_i=1}^{T_i} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(w_{it}-\alpha-\theta_i-\psi_{J(i,t)}-\mu_{i\,J(i,t)}\right)^2}{2\sigma^2} \right] \right)^{m_{it}} \\ \times \prod_{t_i=1}^{T_i-1} \left([\ \lambda_1 \left[1 - G(\psi_{it} + \mu_{it}|\theta_i)\right]]^{c_{it}^J} [\eta]^{c_{it}^f} [\delta]^{c_{it}^m} \right)^{m_{it}} \\ \times \prod_{t_i=1}^{T_i-1} \left([1 - \eta - \delta - \nu - \lambda_1 [1 - G(\psi_{it} + \mu_{it}|\theta_i)]]^{1-c_{it}} \right)^{m_{it}} \\ \times \prod_{t_i=1}^{T_i-1} \left([\lambda_0 (1 - G(b|\psi_i)]^{c_{it}} \left[1 - (\lambda_0(1 - G(b|\theta_i))\right]^{1-c_{it}} \right)^{1-m_{it}} \\ \times \prod_{t_i=1}^{T_i-1} \left[h(\mu_{iN(i,t)}|\theta_i, \psi_{N(i,t)}) f(\psi_{N(i,t)}) \right]^{c_{it}^J} \end{array} \right\} \tag{42}$$

$$\times \prod_{i=1}^{I} \prod_{j=1}^{J} h(\mu_{ij}|\theta_i, \psi_j) s_\psi(\psi_j) s_\theta(\theta_i) \tag{43}$$

The bracketed term contains the likelihood contribution of each individual's work history. The first row in the bracketed term contains the probability of the wage when one is observed. The second and third rows give the probability of differnt kinds of transition for a worker who is currently employed. The fourth row is the probability of exiting or remaining in non-employment for a non-employed worker. The fifth row expresses the probability of moving into a job with an employer of a particular productivity and a match of a particular quality. The final row in the likelihood function contains the probability of the latent heterogeneity, on which the rest of the likelihood function is conditioned.

C Details of the Monte Carlo Simulation Estimates

C.1 Parameters for the Simulated Data

This section presents the parameters used to generate simulated data. The notation is as in the main body of the paper. We use the notation 0 as the employer productivity class label during spells of unemployment. So, the notation δ_{10-} denotes the vector of destination probabilities for a worker with ability type 1 who was unemployed. Note also that the columns of δ are ordered so that the probability of transition to employers of productive type 1 and 2 appear in the first and second columns. The third column is the probability of transition to unemployment.

$$
\begin{aligned}
\pi_A &= (0.50, 0.50) \\
\pi_B &= (0.50, 0.50) \\
\pi_{K|AB} &= \begin{bmatrix} \pi_{K|11} \\ \pi_{K|12} \\ \pi_{K|21} \\ \pi_{K|22} \end{bmatrix} = \begin{bmatrix} 0.50, 0.50 \\ 0.25, 0.75 \\ 0.75, 0.25 \\ 0.50, 0.50 \end{bmatrix} \\
\gamma &= \begin{bmatrix} \gamma_{111} \\ \gamma_{112} \\ \gamma_{121} \\ \gamma_{122} \\ \gamma_{10-} \\ \gamma_{211} \\ \gamma_{212} \\ \gamma_{221} \\ \gamma_{222} \\ \gamma_{20-} \end{bmatrix} = \begin{bmatrix} 0.250 \\ 0.175 \\ 0.100 \\ 0.025 \\ 0.600 \\ 0.150 \\ 0.075 \\ 0.025 \\ 0.025 \\ 0.500 \end{bmatrix} \\
\delta &= \begin{bmatrix} \\ \delta_{111} \\ \delta_{112} \\ \delta_{121} \\ \delta_{122} \\ \delta_{10-} \\ \delta_{211} \\ \delta_{212} \\ \delta_{221} \\ \delta_{222} \\ \delta_{20-} \end{bmatrix} = \left[\begin{array}{cc|c} b=1 & b=2 & b=U \\ \hline 0.4 & 0.1 & 0.5 \\ 0.4 & 0.4 & 0.2 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.5 & 0.2 \\ 0.5 & 0.5 & 0.0 \\ 0.5 & 0.2 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0.5 & 0.4 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.3 & 0.7 & 0.0 \end{array}\right] \\
\theta &= (4, -4) \\
\psi &= (2, -2) \\
\mu &= (1, -1) \\
\alpha &= 0 \\
\sigma &= 0.1
\end{aligned} \tag{44}
$$

C.2 Summary of the Simulated Data

- Match Rates:

$$\begin{bmatrix} m_{K|11} \\ m_{K|12} \\ m_{K|21} \\ m_{K|22} \end{bmatrix} = \begin{bmatrix} 0.55, 0.45 \\ 0.24, 0.76 \\ 0.77, 0.23 \\ 0.54, 0.46 \end{bmatrix}$$

- Separation rates:

$$\begin{bmatrix} s_{111} \\ s_{112} \\ s_{121} \\ s_{122} \\ s_{10-} \\ s_{211} \\ s_{212} \\ s_{221} \\ s_{222} \\ s_{20-} \end{bmatrix} = \begin{bmatrix} 0.270 \\ 0.145 \\ 0.090 \\ 0.026 \\ 0.621 \\ 0.143 \\ 0.092 \\ 0.023 \\ 0.033 \\ 0.578 \end{bmatrix}$$

- Transition Rates:

$$\begin{bmatrix} \\ d_{111} \\ d_{112} \\ d_{121} \\ d_{122} \\ d_{10-} \\ d_{211} \\ d_{212} \\ d_{221} \\ d_{222} \\ d_{20-} \end{bmatrix} = \left[\begin{array}{cc|c} b=1 & b=2 & b=U \\ \hline 0.30 & 0.10 & 0.60 \\ 0.35 & 0.51 & 0.14 \\ 0.83 & 0.00 & 0.17 \\ 0.31 & 0.44 & 0.25 \\ 0.46 & 0.54 & 0.00 \\ 0.49 & 0.20 & 0.31 \\ 0.06 & 0.89 & 0.06 \\ 0.70 & 0.22 & 0.87 \\ 0.09 & 0.82 & 0.09 \\ 0.31 & 0.69 & 0.00 \end{array}\right]$$

C.3 Posterior Mean

$$
\begin{aligned}
\hat{\pi}_A &= (0.55, 0.45) \\
\hat{\pi}_B &= (0.50, 0.50) \\
\hat{\pi}_{K|AB} &= \begin{bmatrix} \hat{\pi}_{K|11} \\ \hat{\pi}_{K|12} \\ \hat{\pi}_{K|21} \\ \hat{\pi}_{K|22} \end{bmatrix} = \begin{bmatrix} 0.55, 0.46 \\ 0.25, 0.75 \\ 0.76, 0.24 \\ 0.56, 0.44 \end{bmatrix} \\
\hat{\gamma} &= \begin{bmatrix} \hat{\gamma}_{111} \\ \hat{\gamma}_{112} \\ \hat{\gamma}_{121} \\ \hat{\gamma}_{122} \\ \hat{\gamma}_{10-} \\ \hat{\gamma}_{211} \\ \hat{\gamma}_{212} \\ \hat{\gamma}_{221} \\ \hat{\gamma}_{222} \\ \hat{\gamma}_{20-} \end{bmatrix} = \begin{bmatrix} 0.27 \\ 0.14 \\ 0.09 \\ 0.03 \\ 0.62 \\ 0.14 \\ 0.09 \\ 0.02 \\ 0.03 \\ 0.58 \end{bmatrix} \\
\hat{\delta} &= \begin{bmatrix} \\ \hat{\delta}_{111} \\ \hat{\delta}_{112} \\ \hat{\delta}_{121} \\ \hat{\delta}_{122} \\ \hat{\delta}_{10-} \\ \hat{\delta}_{211} \\ \hat{\delta}_{212} \\ \hat{\delta}_{221} \\ \hat{\delta}_{222} \\ \hat{\delta}_{20-} \end{bmatrix} = \left[\begin{array}{cc|c} b=1 & b=2 & b=U \\ \hline 0.30 & 0.10 & 0.60 \\ 0.35 & 0.51 & 0.14 \\ 0.83 & 0.00 & 0.17 \\ 0.31 & 0.44 & 0.25 \\ 0.46 & 0.54 & 0.00 \\ 0.49 & 0.20 & 0.31 \\ 0.06 & 0.89 & 0.06 \\ 0.70 & 0.22 & 0.08 \\ 0.09 & 0.82 & 0.09 \\ 0.31 & 0.69 & 0.00 \end{array}\right] \\
\hat{\theta} &= (-4.00, 4.00) \\
\hat{\psi} &= (-2.00, 2.00) \\
\hat{\mu} &= (-1.00, 1.00) \\
\hat{\alpha} &= 1.00 \\
\hat{\sigma} &= 0.10
\end{aligned}
\tag{45}
$$

D Details of LEHD Structural Estimation

D.1 Initial Conditions Using LEHD Data

$$
\begin{aligned}
\pi_A &= (0.50, 0.50) \\
\pi_B &= (0.50, 0.50) \\
\pi_{K|AB} &= \begin{bmatrix} \pi_{K|11} \\ \pi_{K|12} \\ \pi_{K|21} \\ \pi_{K|22} \end{bmatrix} = \begin{bmatrix} 0.50, 0.50 \\ 0.50, 0.50 \\ 0.50, 0.50 \\ 0.50, 0.50 \end{bmatrix} \\
\gamma &= \begin{bmatrix} \gamma_{111} \\ \gamma_{112} \\ \gamma_{121} \\ \gamma_{122} \\ \gamma_{10-} \\ \gamma_{211} \\ \gamma_{212} \\ \gamma_{221} \\ \gamma_{222} \\ \gamma_{20-} \end{bmatrix} = \begin{bmatrix} 0.454572405 \\ 0.395058033 \\ 0.28413216 \\ 0.274265373 \\ 0.30364168 \\ 0.402575865 \\ 0.343973216 \\ 0.278657886 \\ 0.267743695 \\ 0.328570019 \end{bmatrix} \\
\delta &= \begin{bmatrix} \\ \delta_{111} \\ \delta_{112} \\ \delta_{121} \\ \delta_{122} \\ \delta_{10-} \\ \delta_{211} \\ \delta_{212} \\ \delta_{221} \\ \delta_{222} \\ \delta_{20-} \end{bmatrix} = \left[\begin{array}{cc|c} b=1 & b=2 & b=U \\ \hline 0.410297148 & 0.143953654 & 0.445749198 \\ 0.475801944 & 0.205731092 & 0.318466964 \\ 0.231502405 & 0.399017771 & 0.369479824 \\ 0.220494543 & 0.479095915 & 0.300409542 \\ 0.691171279 & 0.308828721 & -- \\ 0.499999527 & 0.198804192 & 0.301196281 \\ 0.492931019 & 0.262196019 & 0.244872962 \\ 0.217817092 & 0.513856712 & 0.268326196 \\ 0.165995755 & 0.572693504 & 0.261310741 \\ 0.61892908 & 0.38107092 & -- \end{array}\right] \\
\theta &= (-0.8810892, 0) \\
\psi &= (-0.67557, 0) \\
\mu &= (-0.5037163, 0) \\
\alpha &= 10.0959986 \\
\sigma &= 1
\end{aligned} \tag{46}
$$

D.2 Posterior Mean in the Structural Estimation Sample

$$
\begin{aligned}
\hat{\pi}_A &= (0.7101, 0.2899) \\
\hat{\pi}_B &= (0.528, 0.472) \\
\hat{\pi}_{K|AB} &= \begin{bmatrix} \hat{\pi}_{K|11} \\ \hat{\pi}_{K|12} \\ \hat{\pi}_{K|21} \\ \hat{\pi}_{K|22} \end{bmatrix} = \begin{bmatrix} 0.2433, 0.7567 \\ 0.9496, 0.0504 \\ 0.4767, 0.5233 \\ 0.9693, 0.0307 \end{bmatrix} \\
\hat{\gamma} &= \begin{bmatrix} \hat{\gamma}_{111} \\ \hat{\gamma}_{112} \\ \hat{\gamma}_{121} \\ \hat{\gamma}_{122} \\ \hat{\gamma}_{10-} \\ \hat{\gamma}_{211} \\ \hat{\gamma}_{212} \\ \hat{\gamma}_{221} \\ \hat{\gamma}_{222} \\ \hat{\gamma}_{20-} \end{bmatrix} = \begin{bmatrix} 0.66033 \\ 0.44177 \\ 0.27959 \\ 0.25058 \\ 0.31277 \\ 0.38144 \\ 0.17986 \\ 0.19232 \\ 0.23731 \\ 0.31785 \end{bmatrix} \\
\hat{\delta} &= \begin{bmatrix} \\ \hat{\delta}_{111} \\ \hat{\delta}_{112} \\ \hat{\delta}_{121} \\ \hat{\delta}_{122} \\ \hat{\delta}_{10-} \\ \hat{\delta}_{211} \\ \hat{\delta}_{212} \\ \hat{\delta}_{221} \\ \hat{\delta}_{222} \\ \hat{\delta}_{20-} \end{bmatrix} = \left[\begin{array}{cc|c} b=1 & b=2 & b=U \\ \hline 0.45464 & 0.11293 & 0.43242 \\ 0.49288 & 0.21597 & 0.29114 \\ 0.2486 & 0.46806 & 0.28335 \\ 0.091005 & 0.28149 & 0.6275 \\ 0.72469 & 0.27531 & -- \\ 0.50792 & 0.031861 & 0.46022 \\ 0.52255 & 0.23242 & 0.24504 \\ 0.11199 & 0.64741 & 0.2406 \\ 0.10885 & 0.62407 & 0.26708 \\ 0.72286 & 0.27704 & -- \end{array}\right] \\
\hat{\theta} &= (-0.2497, 0.6112) \\
\hat{\psi} &= (-1.0961, 1.2256) \\
\hat{\mu} &= (-0.7243, 1.0562) \\
\hat{\alpha} &= 9.0777 \\
\hat{\sigma} &= 0.4330
\end{aligned} \tag{47}
$$