

Modeling Endogenous Mobility in Wage Determination

John M. Abowd Department of Economics, Cornell University,
 john.abowd@cornell.edu
Kevin L. McKinney Center for Economic Studies, U.S. Census Bureau,
 kevinm@ccrdc.ucla.edu
Ian M. Schmutte Department of Economics, University of Georgia,
 schmutte@uga.edu

May 2015

Abstract

We evaluate the bias from endogenous job mobility in fixed-effects estimates of worker- and firm-specific earnings heterogeneity using longitudinally linked employer-employee data from the LEHD infrastructure file system of the U.S. Census Bureau. First, we propose two new residual diagnostic tests of the assumption that mobility is exogenous to unmodeled determinants of earnings. Both tests reject exogenous mobility. We relax the exogenous mobility assumptions by modeling the evolution of the matched data as an evolving bipartite graph using a Bayesian latent class framework. Our results suggest that endogenous mobility biases estimated firm effects toward zero. To assess validity, we match our estimates of the wage components to out-of-sample estimates of revenue per worker. The corrected estimates attribute much more of the variation in revenue per worker to variation in match quality and worker quality than the uncorrected estimates.

Keywords: Earnings heterogeneity, Mobility Bias, Latent Class Model,
Markov Chain Monte Carlo

1 Introduction

We study the consequences of endogenous mobility for estimates of worker and firm effects on labor market earnings. Our starting point is the fixed effects estimator for the statistical earnings model developed by Abowd et al. (1999, AKM, henceforth), which decomposes log earnings into components associated with unobserved worker and employer heterogeneity. Their statistical model requires the assumption that mobility is exogenous to the earnings residual, or, equivalently, that the assignment of workers to firms is random conditional on all observable characteristics and the time-invariant unobservables. Structural interpretations of the estimated worker and firm effects are potentially useful as measures of unobserved skill and firm-specific wage premia as long as mobility is exogenous. However, the nature of any endogenous mobility bias is necessarily complex and difficult to assess. Our goal is to determine whether relaxing the exogenous mobility assumption has major implications for estimates of worker and employer-specific earnings heterogeneity.

We begin by clearly defining the exogenous mobility assumption entailed by the log earnings model proposed in AKM. The assumption of exogenous mobility implies that job mobility and job assignment depend only on time-invariant unobservable characteristics of workers and firms. While this allows for certain types of sorting, it precludes mobility driven by learning about new outside job opportunities that carry match-specific earnings premia (Woodcock 2008). It also rules out models in which mobility is associated with learning about comparative advantage (Gibbons et al. 2005), or idiosyncratic labor demand shocks (Helwege 1992). The exogenous mobility assumption is, therefore, subject to considerable skepticism among economists.

Despite the widespread skepticism that mobility is endogenous to the wage residual, there are no tests of exogenous mobility that can be computed from estimates made under the null hypothesis. We fill this gap by developing two new residual diagnostic tests of the exogenous mobility assumption, which we apply to longitudinally-linked employer-employee data from the U.S. Census Bureau's Longitudinal Employer Household Dynamics (LEHD) program. In both tests, the LEHD data reject the null hypothesis of exogenous mobility.

Rejecting the exogenous mobility assumption leaves open the question of whether and how endogenous mobility biases the estimates of worker- and firm-specific contributions to pay. To

address these questions, we develop a latent class model of the data generating process that relaxes two key assumptions of the exogenous mobility model. Our model allows, first, for match effects that are correlated with worker and firm heterogeneity. Second, we allow both earnings and job mobility to be determined by latent classifications of workers, firms, and matches. We estimate the model using Bayesian methods that, because of the non-nested structure of the latent classifications, become computationally intensive. We exploit the network structure of the data and associated model restrictions to speed computation by using a graph coloring algorithm.

We show that the uncorrected estimates from the AKM model derived under the assumption of exogenous mobility support our modeling decisions. The results from estimating the endogenous mobility model suggest that (1) match effects are a substantial determinant of earnings heterogeneity; (2) firm effects are biased toward zero in the uncorrected model; (3) there is very little structural correlation between worker and firm components of earnings heterogeneity; (4) match-specific wage variation acts to compress, rather than amplify, differences in earnings within and across firms.

We go on to validate our estimates using data on firm revenues. Using the uncorrected AKM estimates, we find a strong association of firm revenue with firm effects, but only a very weak association with firm-average worker and firm-average match effects. Using the endogenous-mobility corrected estimates, we find revenue is strongly associated with all three components of wage heterogeneity. The latter pattern is consistent with models in which worker and match effects in earnings represent compensation for productive attributes.

Our analysis and methods are relevant in applications that use estimates of worker and firm effects as measures of human capital and firm-specific wage premia. Examples include the analysis of firm productivity (Iranzo et al. 2008), inter-industry wage differentials (Abowd et al. 2012), exporter wage differentials (Krishna et al. 2014), job referral networks (Schmutte 2015), and wage inequality (Card et al. 2013). Endogenous assignment also affects related empirical settings, including the estimation of neighborhood effects on earnings (Combes et al. 2008) and the effects of teachers on student test scores (Rothstein 2010; Koedel and Betts 2010). Kramarz et al. (2014) use policy-induced random variation to correct for endogenous school assignment. They find that endogenous assignment biases estimated school effects toward zero. We find a very similar result,

albeit using rather different data and empirical methods.

2 Background and Motivation

2.1 The AKM Model for Labor Market Earnings

AKM proposed a framework for modeling individual log wages in matched employer-employee data that allows for arbitrary heterogeneity across workers and across employers. In the AKM model, the log wage of worker i in period t is

$$\ln w_{it} = X_{it}\beta + \theta_i + \psi_{J(i,t)} + \varepsilon_{it}. \quad (1)$$

The log wage depends on observable characteristics through $X_{it}\beta$, unobservable individual-level characteristics, θ_i , and an unobservable employer-specific component, $\psi_{J(i,t)}$, where $J(i,t)$ is a function that maps worker-year observations to their unique employer in that year. This model applies to data that include observations on I individual workers and J employers. The data are observed over T periods and in any period t , there are a total of N_t observations. The full sample includes N worker-period observations.

In matrix notation, the wage model (1) is

$$\ln w = X\beta + D\theta + F\psi + \varepsilon \quad (2)$$

where $\ln w$ is the $[N \times 1]$ stacked vector of log wage outcomes $\ln w_{it}$, X is the $[N \times k]$ design matrix of observable time-varying characteristics. In the analysis that follows, and in the empirical work, $\ln w$ and X are measured as deviations from their overall means and we suppress the constant term. The matrix D is the $[N \times I]$ design for the individual effects; F is the $[N \times J]$ design matrix for the employer effects (non-employment is suppressed here). The unknown effects to be estimated, $\left[\beta^T \quad \theta^T \quad \psi^T \right]^T$, have dimension $[k \times 1]$, $[I \times 1]$, and $[J \times 1]$ associated with each of the design matrices.

Identification and the Exogenous Mobility Assumption

Identification of the parameter vector $\begin{bmatrix} \beta^T & \theta^T & \psi^T \end{bmatrix}^T$ in the statistical model requires the following orthogonality conditions:

$$\text{E} [X^T \varepsilon] = 0; \text{E} [D^T \varepsilon] = 0; \text{E} [F^T \varepsilon] = 0. \quad (3)$$

As long as the matrix of data moments has full rank, this conditional moment restriction yields a consistent estimator for the full parameter vector, including the individual and employer effects.

The assumptions in (3) that $\text{E} [D^T \varepsilon] = 0$ and $\text{E} [F^T \varepsilon] = 0$ are particularly problematic. They imply that there is no correlation between the wage residual and an individual's decision to enter or exit the labor market, and that there is no correlation between the residual and the assignment of workers to employers. These assumptions do not have a clear behavioral foundation, but they follow from the stronger assumption that mobility and assignment are independent of the wage residual.

Specifically, we define exogenous mobility by the assumptions:

$$\begin{aligned} \text{E} [\varepsilon|X] &= 0 \\ \text{Pr} [D, F|X, \varepsilon] &= \text{Pr} [D, F|X]. \end{aligned} \quad (4)$$

If the exogenous mobility assumptions are satisfied, it is clear that $\text{E} [\varepsilon|X, D, F] = 0$. The orthogonality conditions necessary for identification follow. See Abowd et al. (2002) for the method of ensuring that D and F have full column rank.

Exogenous mobility requires that a worker's employment history is completely independent of the idiosyncratic part of earnings captured in ε . Specifically, knowledge of the entire history of wage residuals does not convey any information that would help predict job assignment or worker entry and exit. The exogenous mobility assumption is, thus, equivalent to assuming that all assignments are pre-determined at birth given full knowledge of X, D, F and $\begin{bmatrix} \beta^T & \theta^T & \psi^T \end{bmatrix}^T$ (Rothstein 2010).

2.2 Testing the Exogenous Mobility Assumption

We formulate two tests that exploit the exogenous mobility assumption that current earnings residuals should not be predictive of future employer assignments, and apply them to longitudinally integrated employer-employee data from the Longitudinal Employer-Household Dynamics (LEHD) Program of the U.S. Census Bureau. Both tests strongly reject the null hypothesis of exogenous mobility for these data. We describe the data and procedure for calculating both tests in detail in Appendix A.

The first test, the *match effects test*, checks whether the firm effect of a worker's future employers are independent of the average residual in the current job. The match effects test yields a test statistics, $X^2 = 7,438,692$, that is distributed chi-square with 8,991 degrees of freedom. Using conventional criteria, this test has a p-value less than 10^{-6} . From a Bayesian viewpoint and assuming equal prior odds on the null and alternative, the change in the Bayes Information Criterion (BIC) associated with going from the null hypothesis to the unconstrained alternative hypothesis is 166,049, which also indicates that the data strongly favor some model in which the AKM residuals are related to job mobility.

The second test, the *productive workforce test*, checks whether the average worker effect of future employees of a particular employer are predictive of the residuals on that firm's current period wage payments. The productive workforce test has a test statistic $X^2 = 172,295$ that is distributed chi-square with 900 degrees of freedom. Again, using conventional criteria, this test has a p-value less than 10^{-6} . The change in the BIC is 13,714, which also strongly favors some model in which the AKM residuals are related to job mobility.

3 Empirical Model

To relax the assumption of exogenous mobility, we develop a model in which a match-specific component of the earnings residual predicts the movement of workers between employers. The huge number of workers and employers render the problem of predicting job mobility extremely challenging. To make progress, we use a latent class framework. Each of the three populations of interest – workers, employers, and job matches – are associated with latent heterogeneity classes

that affect wages, mobility and job assignment. Our model allows for arbitrary correlation between worker and employer types on observed matches, and also allows for sorting by comparative advantage by allowing arbitrary correlation between job match quality and worker and employer attributes.

3.1 Model Setup

Population Heterogeneity

The agents are workers, indexed $i \in \{1 \dots I\} \equiv I$ and employers, indexed $j \in \{0 \dots J\} \equiv J$. By convention, when a worker is not employed, we say he is assigned to employer $j = 0$. On entry to the labor market, a worker i samples his type from one of L latent ability classes, $a_i \in A$. Likewise, each employer, except $j = 0$, samples her type from one of M latent classes, denoted $b_j \in B$. Again, by convention, the non-employment state is associated with its own latent class, $b_0 = M + 1$

In our empirical application, we have access to data on the complete population of workers and employers. Each potential worker-employer match (a job) has an associated latent heterogeneity component that affects both wages and mobility: $k_{ij} \in K$, where K has cardinality Q . To make the subsequent formulas easier to interpret, we represent the elements of A , B and K as rows from the identity matrices I_L , I_{M+1} and I_Q , respectively. For instance, if the number of latent worker types, $L = 2$, $A = \{(1, 0), (0, 1)\}$.

We assume workers and employers sample their latent ability and productivity classes independently from multinomial distributions with parameters π_a, π_b . However, the distribution of match quality is not independent of worker or employer type. The probability that the latent quality of the match between worker i and employer j is of type k is $\Pr(k_{ij} = k | a_i = a, b_j = b) = \pi_{k|ab}$. This specification allows for independent match effects as a special case. If match quality is independent of worker and employer heterogeneity, then AKM estimates of the worker and firm effects are unbiased even in the presence of endogenous mobility.

Earnings Determination

The log of earnings on any match is given by the following generalization of the AKM model

$$\ln w_{ijt} = \alpha + X_{ijt}\beta + a_i\theta + b_j\psi + k_{ij}\mu + \varepsilon_{it} \quad (5)$$

where X_{ijt} is a vector of observable time-varying characteristics and 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.

Mobility Model

We relax exogenous mobility by allowing those matches and employment durations that are observed to depend on match quality. The separation indicator, $s_{it} = 1$ if i separates from his current job at the end of period t , and $s_{it} = 0$ otherwise. Recall $J(i, t)$ is the index function that returns the identifier of the firm in which i is employed in period t . The probability of separation depends flexibly on match quality:

$$\Pr [s_{it} = 1 | k_{iJ(i,t)}] = f_{se} (a_i, b_{J(i,t)}, k_{iJ(i,t)}; \gamma) \equiv \gamma_{abk} \quad (6)$$

where $0 \leq \gamma_{abk} \leq 1$.

Furthermore, conditional on separation, the productivity class of the next employer depends on the quality of the current match, in addition to the productivity of the current employer and the ability of the worker:

$$\Pr [b_{J(i,t+1)} | a_i, b_{J(i,t)}, k_{iJ(i,t)}] = f_{tr} (a_i, b_{J(i,t)}, k_{iJ(i,t)}; \delta) \equiv \delta_{abk} \in \Delta^{M+1} \quad (7)$$

where $\delta_{abk} \equiv [\delta_{0|abk}, \dots, \delta_{M|abk}]$ is a $1 \times (M + 1)$ vector of transition probabilities, Δ^{M+1} is the unit simplex, and $J(i, 0) = 0$ 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.

3.2 The Network Interpretation of Labor Mobility

The labor market is an evolving graph of connections between workers and their employers. At time t , let the set of identifiers for all I individuals who work in one of the J employers (including non-employment), $P(t)$, 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 the links (jobs) active at date t can be represented as an $I \times J$ matrix $B(t)$, which is the upper right-hand block of the full adjacency matrix for the bipartite graph, with i on the rows and j on the columns, and is sometimes called the biadjacency matrix.

The collection of labor market relationships at time t , summarized by the adjacency matrix $B(t)$, we call the *realized employment network*. The observed labor market data are snapshots of the market at points in time, $B(t_1), \dots, B(t_T)$, where T is the total number of available time periods. Our adjacency matrix representation can be directly related to the AKM framework. When the data are sorted first by time, t , and then by workers, i , the design matrix of employer effects is

$$F = \begin{bmatrix} B(1) \\ B(2) \\ \vdots \\ B(T) \end{bmatrix}$$

where $B(t)$ is the adjacency matrix from the bipartite labor market graph.

Each adjacency matrix, $B(t)$, describes which outcomes (job matches) were observed at each point in time from the collection of $I \times J + 1$ potential outcomes at each moment of time. The potential outcomes are given by a structural wage equation, which we introduced in equation (5). In this sense, the problem of endogenous mobility is a sample selection problem. Under exogenous mobility, potential outcomes are selected for observation conditionally at random given X , D , and F .

To address these selection biases, our model groups together workers and firms with similar mobility and earnings patterns by assigning them to latent types. Our application and proposed procedure are, therefore, related to stochastic block models, modularity maximization, and other

methods for the detection of “communities” of nodes in complex social and economic 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; Schmutte 2014).

3.3 Likelihood Function

The observed data, y_{it} , consist of wage rates, observable time-varying characteristics, separations, accessions, and identifier information:

$$y_{it} = [\ln w_{iJ(i,t)t}, X_{it}, s_{it}, i, J(i, t), J(i, t + 1)] \text{ for } i = 1, \dots, I \text{ and } t = 1, \dots, T. \quad (8)$$

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

$$Z = [a_1, \dots, a_I, b_1, \dots, b_J, k_{11}, k_{12}, \dots, k_{1J}, k_{21}, \dots, k_{IJ}]. \quad (9)$$

Finally, the complete parameter vector is

$$\rho^T = [\alpha, \beta^T, \theta^T, \psi^T, \mu^T, \sigma, \gamma, \delta, \pi_a, \pi_b, \pi_{k|ab}], \rho \in \Theta. \quad (10)$$

The observed data matrix is denoted Y . The likelihood function for the parameters is the joint distribution of observed and latent data (Y, Z) :

$$\begin{aligned} \mathcal{L}(\rho|Y, Z) \propto & \prod_{i=1}^I \left\{ \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(\ln w_{it} - \alpha - X_{it}\beta - a_i\theta - b_{J(i,t)}\psi - k_{iJ(i,t)}\mu)^2}{2\sigma^2} \right] \right. \\ & \times \prod_{t=1}^{T-1} \left[1 - \gamma_{\langle a_i \rangle} \langle b_{J(i,t)} \rangle \langle k_{iJ(i,t)} \rangle \right]^{1-s_{it}} \left[\gamma_{\langle a_i \rangle} \langle b_{J(i,t)} \rangle \langle k_{iJ(i,t)} \rangle \right]^{s_{it}} \\ & \times \prod_{t=1}^{T-1} \left[\delta_{\langle b_{J(i,t+1)} \rangle} | \langle a_i \rangle \langle b_{J(i,t)} \rangle \langle k_{iJ(i,t)} \rangle \right]^{s_{it}} \left. \right\} \\ & \times \prod_{i=1}^I \prod_{j=1}^J \prod_{\ell=1}^L \prod_{m=1}^M \prod_{q=1}^Q \left[\left(\prod_{\ell=1}^L \prod_{m=1}^M \prod_{q=1}^Q (\pi_{a\ell})^{a_{i\ell}} (\pi_{bm})^{b_{jm}} (\pi_{q|\ell m})^{k_{ijq}} \right) \right] \end{aligned} \quad (11)$$

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 indicator vector x .

The likelihood function 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. We assume that the matches initially observed are exogenous.

The power of the model comes from the predictive equation for the latent data Z – the unobserved types associated with worker, employer, and match heterogeneity. Given the observed data and the parameters, the posterior predictive distribution of Z is computed as the complete-data likelihood divided by the observed-data likelihood. The observed-data likelihood is calculated by integrating out the latent data. We describe our estimation procedure in detail in Section 4.

4 Estimation Method

Our empirical approach is Bayesian, but standard estimation techniques are not effective because the worker and employer effects are not nested. We estimate the model by adapting the Gibbs sampler for finite mixture models as developed in Tanner (1996) and Diebolt and Robert (1994) to our model, allowing for multiple overlapping levels of correlation across observations. As in the data augmentation algorithm, we iterate between sampling from the posterior distribution of the parameters given the complete data – both observed and latent data – and sampling from the posterior predictive distribution of the latent data – the unobserved worker, employer, and match types – given the model parameters.

Below, we derive the posterior distribution of the model parameters given the latent data. The derivation is standard, as is the method for sampling. The more challenging task, both analytically and computationally, is deriving, and sampling from, the posterior predictive distribution for unobserved types. Under the model, the posterior distribution for the latent type depends on the relative likelihood contribution of a particular entity (worker, employer, or match) under each latent classification. Generically, the likelihood contribution of a worker depends not only on his own ability classification, but also on the productivity classifications of the employers he works for, and the

match quality.

We exploit the relational structure of the data together with conditional independence assumptions implied by the model to facilitate computation. We can update the worker effects first, then the employer effects, then the match effects. Furthermore, the conditional independence assumptions in the model imply the latent classifications of the workers can be updated simultaneously (in parallel). The same is true of the match heterogeneity classifications.

For employer heterogeneity, the situation is more complex. Because the probability of assignment to a new employer is a function of the previous employer, the likelihood contribution of a given employer is not independent of the employers to which it is directly connected through the realized mobility network. Therefore, the latent classification of any employer is not independent of the classification of its network neighbors. Without further analysis, this requires the employer classifications be updated sequentially, which is very time consuming.

We use the network structure of the data to parallelize the computation. Specifically, the employer types are independent across groups of firms that are not connected through a direct job-to-job transition; that is, firms that do not have a degree-one network connection. We apply a graph coloring algorithm to the employer projection of the realized mobility network to partition employer nodes into disjoint groups, based on this conditional independence, that can be updated in parallel.

4.1 Prior and Posterior distributions

The prior on the parameter vector $\rho^T = [\alpha, \beta^T, \theta^T, \psi^T, \mu^T, \sigma, \gamma, \delta, \pi_a, \pi_b, \pi_{k|ab}]$ is the product of priors on its component terms. Each vector of probabilities, $\gamma, \delta, \pi_a, \pi_b, \pi_{k|ab}$, has a Dirichlet prior. Each element of the Dirichlet parameter given by the inverse of the dimension of the probability vector.

We choose conjugate priors for the wage and mobility model parameters. Conditional on the population type probabilities, $\pi_a, \pi_b, \pi_{k|ab}$, the coefficients in the log wage equation, $(\alpha, \beta^T, \theta^T, \psi^T, \mu^T)$ have uninformative normal prior distributions. The variance parameter, σ , has the inverted gamma prior $IG(\nu_0, s_0)$. In practice, we set $\nu_0 = 1$ and $s_0 = 1$.

We also constrain the population probability-weighted average earnings heterogeneity effects

to be zero. That is, $\pi_a^T \theta = \pi_b^T \psi = \pi_{k|ab(\ell m)}^T \mu = 0$ for all ℓ, m where $\pi_{k|ab(\ell m)} \equiv \Pr[k_{ij} = k | a_i = \ell, b_j = m]$. This assumption highlights the inherent sample selection problem. If all workers were observed in all matches in all periods, there would be no endogenous mobility bias. With the preceding assumptions, we derive the posterior distributions for each parameter in Appendix B.

4.2 Gibbs Sampler

We start the Gibbs sampler with initial values for the parameter vector and latent data, $\rho^{(0)}, Z^{(0)}$. To update the parameter vector, we sample from the posterior distributions defined in Appendix B. We must still define the posterior distribution for the latent data, Z , given the observed data and parameters. To update the ability classifications for the workers, we sample from a multinomial with probability that worker i falls in the ℓ^{th} class:

$$\begin{aligned} \Pr(a_i = \ell | a_{-i}, b, k, Y, \rho) &= \frac{\Pr(a_{-i}, b, k, Y | \rho, a_i = \ell) \Pr(a_i = \ell)}{\Pr(a_{-i}, b, k, Y | \rho)} \\ &= \frac{\pi_{a\ell} \Pr(a_{-i}, b, k, Y | \rho, a_i = \ell)}{\sum_{\ell'=1}^L [\pi_{a\ell'} \Pr(a_{-i}, b, k, Y | \rho, a_i = \ell')]} \end{aligned} \quad (12)$$

where a_{-i} represents the classifications of all workers other than i . To calculate Equation 12 requires computing the likelihood contribution of i when assigned to each of the L ability classifications. Under our model, it is easy to show the posterior probability of a_i is independent of a_{-i} , conditional on the rest of the data (latent and observed),

$$\Pr(a_i = \ell | a_{-i}, b, k, Y, \rho) = \Pr(a_i = \ell | b, k, Y, \rho). \quad (13)$$

This conditional independence allows us to speed computation by updating the latent classifications of each worker in parallel. The proof is a straightforward consequence of the conditional independence across workers in the likelihood function for the complete data.

The posterior predictive distribution for the latent match quality is

$$\Pr(k_s = q | a, b, k_{-s}, Y, \rho) = \Pr(k_s = q | a, b, Y, \rho), \quad (14)$$

which likewise follows from the conditional independence assumptions in the model. Hence, for a given classification of workers and employers, (a, b) , the latent quality of each match is conditionally independent from the others. We exploit the conditional independence by parallelizing these updates as well.

The posterior distribution for employer types exhibits a conditional dependence that is not present for workers or matches. When a worker changes jobs, the latent type of the employer for the successor job depends on the the latent type of the employer on the origin job. Therefore, the posterior probability that a firm is of a particular type depends directly on the types of firms it is connected to through the realized mobility network.

The posterior distribution is

$$\Pr(b_j = m | a, b_{-j}, k, Y, \rho) = \Pr(b_j = m | a, b_{N(j)}, k, Y, \rho) \quad (15)$$

where $b_{N(j)}$ denotes the latent classifications of the employers in $N(j)$, the set of neighbors of j (in the employer projection of the realized mobility network). We use this result to define a partition of employers into groups that can be updated in parallel. The details of our graph coloring algorithm appear in [Appendix D.1](#).

5 Results

5.1 Data

We implement the model empirically using matched employer-employee data from the LEHD program of the U.S. Census Bureau. Our analysis compares estimates from the standard AKM decomposition applied to the entire sample with estimates from the model described in [Section 3](#). The comparison indicates that relaxing the exogenous mobility assumption may have a large effect on estimated worker, employer, and match effects. To validate our model, we bring in data on firm revenue, which is not part of the LEHD infrastructure file system. This section briefly describes the sources of data and how we prepared the research files.

The LEHD Data and AKM Decomposition

We begin with an analysis of the AKM decomposition using the full universe of LEHD dominant job records over the period 1990–2010. The basic structure of the LEHD data is described in Abowd et al. (2009). The data processing and procedures used to estimate the AKM decomposition for these data are almost exactly as described in Abowd et al. (2003). The key difference is that in our case, the AKM model was estimated on the full, national, sample, whereas Abowd et al. (2003) estimated the model state-by-state and then combined the state-level estimates.

We also use the population estimates of the AKM parameters, θ_{AKM} and ψ_{AKM} to create initial classifications of workers, firms, and matches for the Gibbs sampler. For each match, we construct an AKM orthogonal match effect, μ_{AKM} , as the average residual during the match. Next, we construct the deciles of θ_{AKM} , ψ_{AKM} , and μ_{AKM} within their respective populations (across workers, across firms, and across matches). We initialize the Gibbs sampler by assigning each worker, each firm, and each match to its corresponding decile. We discuss these starting values, which also have implications for the exogenous mobility assumption, in Section 5.2.

Data for Structural Estimation

For estimation of our structural model, we restrict analysis to the population of workers who were employed in the states of Illinois, Indiana, and Wisconsin in the in the years 1999–2003. There are 16.9 million workers in this three-state sub-population. For each of these workers, we attach all dominant jobs in which they were ever employed, regardless of the state in which the employment occurred, over the years 1990–2010. Thus, we include the complete 21-year dominant job work history for each worker.

Estimation of the structural model on the full subpopulation is computationally infeasible. We therefore draw a 0.5% simple random sample of workers from the analysis population, retaining their full employment histories. The final analysis sample includes 1,778,490 Person-year observations (including years spent in non-employment) that cover 84,690 Persons, 181,592 Firms, and 389,718 Matches. We assign each observation in the analysis sample the appropriate θ -decile, ψ -decile, and μ -decile based on the AKM estimates to use as starting values for the latent classification of workers, employers, and matches.

Firm Revenue Data

We match our analysis sample with firm-level data on total revenue per worker. The firm-level revenue data is built up from the 2002 Economic Census. The population for the revenue data is all establishments that appear in the Census Business Register. The Economic Census collects sales data from a sample of establishments. The sampling is based on industry and establishment size. For non-sampled establishments, which are missing conditionally at random, missing revenue and capital data is multiply imputed using a model that conditions on all frame variables. The results is a dataset that contains the universe of all business establishments with complete data on revenue and employment. Next, the establishments in the LEHD are matched to establishments in the Business Register. The establishment identifiers in the LEHD are distinct from the establishment identifiers in the Census Business Register, necessitating the use of a statistical matching procedure. For multi-unit firms, we sum revenue and employment across all establishments and compute revenue per worker directly.

5.2 Diagnostic Analysis of the AKM Estimates

Evidence for Match Effects in the AKM Results

We begin by presenting diagnostic statistics based on the AKM model estimated under the assumption of exogenous mobility. Figures 1 and 2 display the average (orthogonal) match effect (μ_{AKM}) within cells defined by unique pairs of worker effect - firm effect deciles. Figure 1 displays this information as a three-dimensional bar graph, akin to the presentation in Abowd et al. (2010) and Card et al. (2013). Unlike the German data analyzed by Card et al. (2013), the LEHD sample exhibits more variation in the average match effect across cells. Therefore, we report the same information in Figure 2 as a grouped bar graph. The data are grouped by worker effect deciles along the horizontal axis. Within each group are plotted bars for each firm effect decile, given the worker effect decile, whose heights correspond to the average match effect within each. Figure 2 is thus a “flattened” version of Figure 1. While the average match effects are less than 2 log points in many cells, their magnitudes are sometimes quite large and vary considerably across worker deciles and across firm deciles. In particular, low wage workers seem to be employed in

high-paying matches. Conversely, there is some tendency for high wage firms to employ workers on low-paying matches, though this pattern is not as strong. As a diagnostic tool, Figure 1 supports our inclusion of correlated match effects in the structural model.

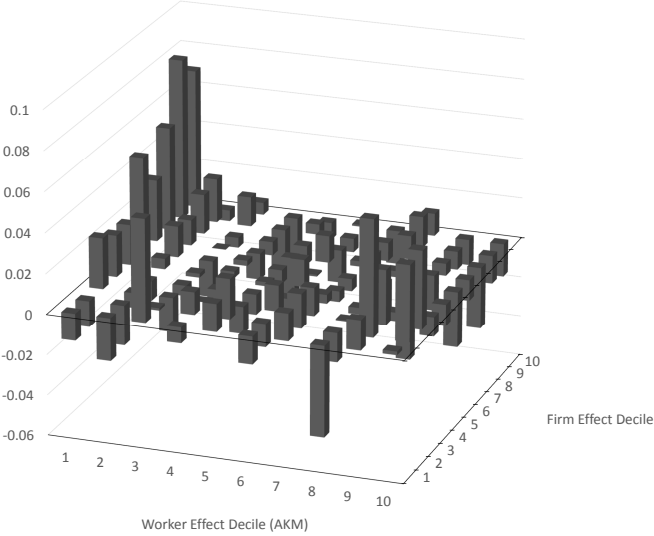


Figure 1: Mean residual from the AKM decomposition by Worker/Firm effect deciles.

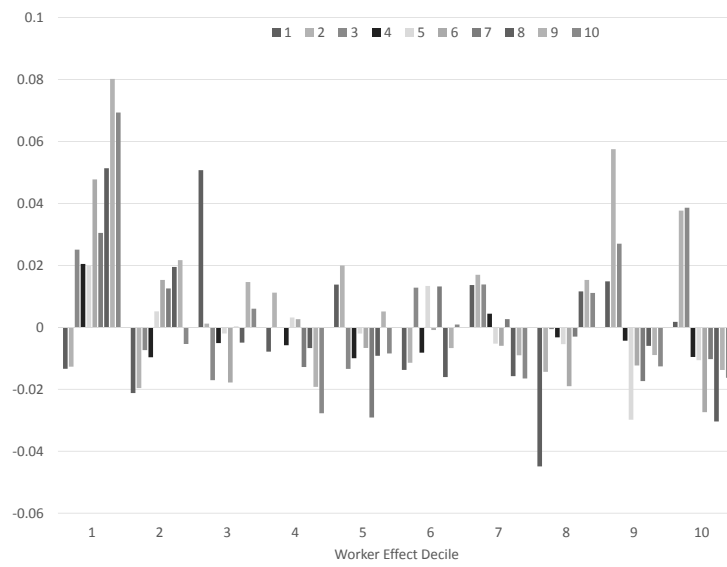


Figure 2: Mean residual from the AKM decomposition by Worker/Firm effect deciles, grouped bar graph. Legend is firm types.

Evidence on the Exogenous Mobility Assumption

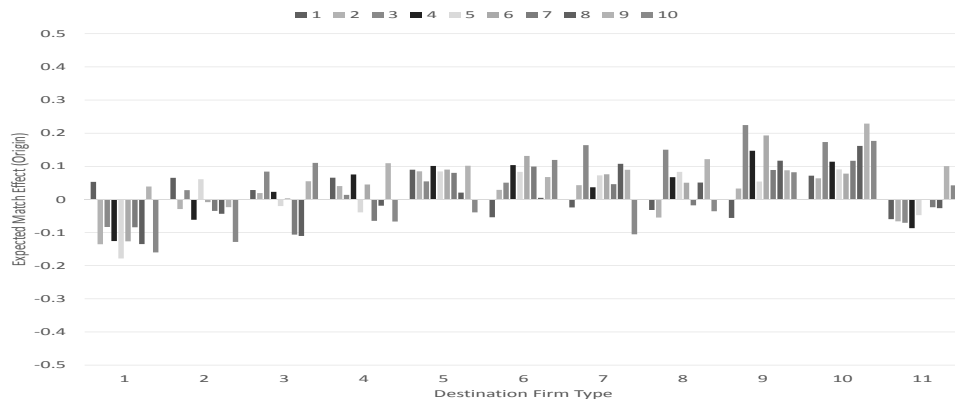
The exogenous mobility assumption in Equation 4 implies, among other things, that for workers who change jobs, the match effect (average residual) on the origin job should not help predict the identity of the destination job. A straightforward corollary, following Bayes' Rule, is that the average match effect for workers changing jobs is independent of the type of firm to which they are transitioning. Knowledge of the future firm type is uninformative about the current match effect after controlling for the type of the worker and of the origin firm.

We assess this corollary implication graphically in Figure 3. The figure displays three grouped bar graphs that each display, for a fixed worker type, the average match effect within cells defined by origin firm–destination firm pairs. For parsimony, we display results for three worker types – those in the first, fifth, and tenth decile of the AKM worker effect distribution. Each sub-figure is organized with destination firm types along the horizontal axis. Within each destination firm type are displayed 10 bars that each correspond to the average match effect for a specific origin firm type.

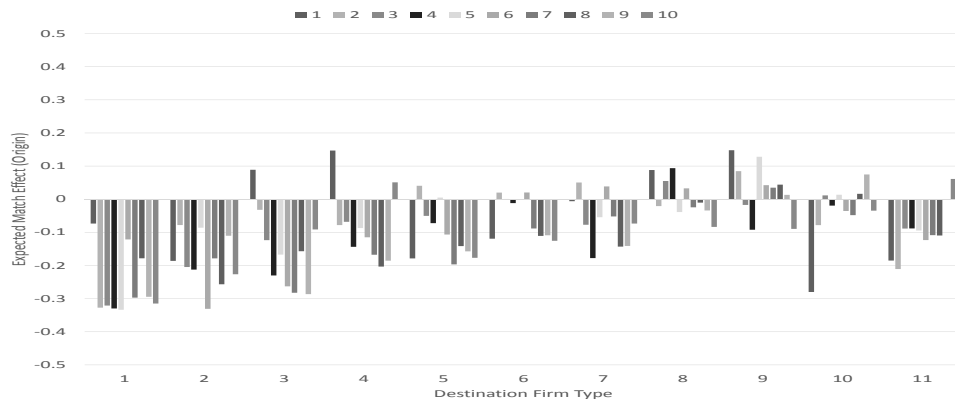
If the assumption of exogenous mobility were valid, the pattern of average match effects should not vary across destination firm types. Graphically, Figure 3a, for instance, would appear as a repeating pattern across all destination firm types. Instead, there is a visually evident correlation between destination firm and the average match effect. Specifically, it appears that movements to low wage firms are predicted by low-wage matches. For low wage workers (Worker Type 1), high average match effects on the origin job are predictive of movements to higher wage firms, conditional on separation. Recall that we code transitions to non-employment as firm type “11”. For all worker types, transitions to non-employment are predicted by low average match effects on the origin job.

Finally, the data show some evidence of wage compression across worker types, and across firms within worker types. The average match effects are larger, and more negative, as we move from Worker Type 1 (low wage workers) to Worker Type 10 (high wage workers). Focusing on Figure 3c, within a specific destination firm type, there is some evidence that the average match quality is higher on origin jobs in low-wage firms and lower on origin jobs in high-wage firms. The same pattern is exhibited across firm types in Figure 2. Altogether, these plots cast doubt

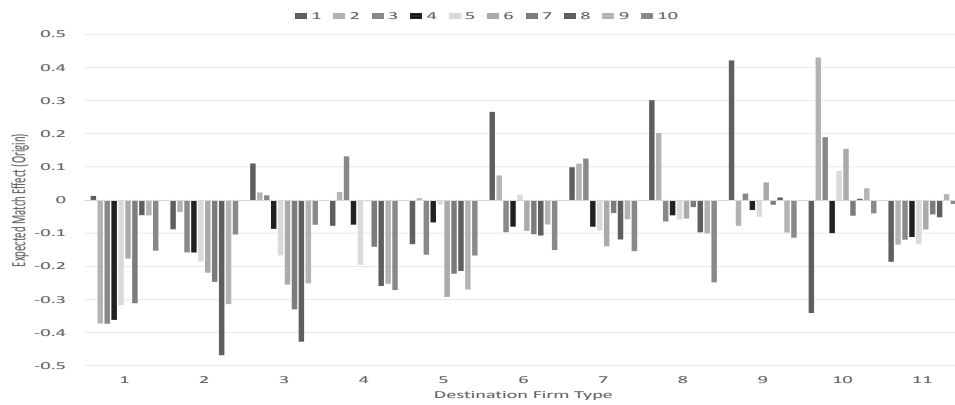
on the plausibility of the exogenous mobility assumption for the LEHD data, consistent with the residual diagnostic tests in Section 2.2. Furthermore, the compression pattern in the data suggests endogenous mobility may bias estimated worker and firm effects toward zero.



(a) Worker Effect Decile 1



(b) Worker Effect Decile 5



(c) Worker Effect Decile 10

Figure 3: Mean AKM residual within origin/destination firm effect decile, disaggregated by worker effect decile. Legend is firm types.

5.3 Estimation Details

We fit the structural model to the 0.5% LEHD analysis sample using the Gibbs sampler. Our results are based on 7,922 draws, taken in approximately equal proportion, from three parallel runs of the Gibbs sampler. The sampler appears to converge after roughly 300 iterations, but exhibits extensive within-chain autocorrelation. Because of the computational demands, we could not take a large enough sample to eliminate the effects of autocorrelation by thinning the sample by, say, selecting every 1,000th sample to analyze. Instead, we report Monte Carlo standard errors, as described in Appendix Section D.2, that properly account for the serial correlation within each sequence of draws from the sampler.

Before the final estimation, we engaged in a model selection step to determine the number of support points for the distribution of latent person, employer, and match heterogeneity. We base model selection on the criterion that the structural model should explain as much of the variation in earnings as the AKM decomposition. We estimated the model, adding support points to latent heterogeneity distributions until the variance of the structural earnings residual was near the variance of the AKM residual (after removing the orthogonal AKM match effect). That is, we sought a model with granularity sufficient to have as much explanatory power in the wage equation as the unrestricted AKM model.

We report results based on a model in which there are ten points of support for each distribution. That is, there are ten worker types ($L = 10$), ten employer types ($M = 10$), and ten match types ($Q = 10$). A comparison of the correlation with log earnings of ε_{Gibbs} and ε_{AKM} reported in Table 2 shows we were able to obtain a good fit on residual variation. The correlation between the AKM residual and log earnings is $\text{Corr}(y, \varepsilon_{AKM}) = 0.2003$, while the correlation between the structural residual and log earnings is $\text{Corr}(y, \varepsilon_{Gibbs}) = 0.2687$. In practice, our model may include more support points for the employer heterogeneity than needed. If so, the apparent lack of parsimony does not introduce problems with our results, since we can collapse redundant classes in our over-parameterized model.

5.4 Summary of Structural Earnings Model Estimates

Figure 4 depicts the posterior distribution of the structural wage equation parameters. The figures plot the posterior mean together with the 5th and 95th percentile of the posterior distribution (left column) and the posterior mean $\pm 2 \times MCSE$ (right column). For consistency with the conventional AKM decomposition, we report the wage parameters as deviations from the grand mean of log earnings.

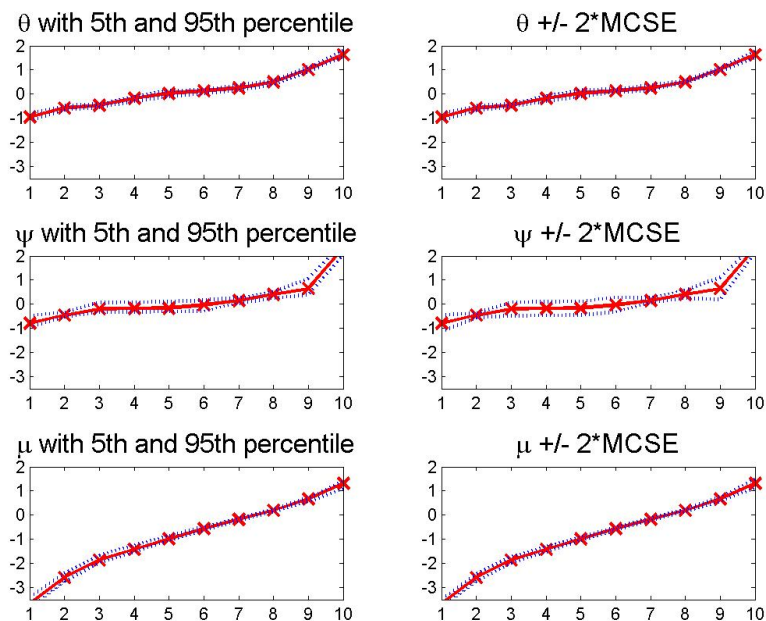


Figure 4: Posterior distribution of wage equation parameters

Note that there is variation in the estimated earnings parameters on all three heterogeneity dimensions. The dispersion of the match effects is much greater than that of the person effects (θ) or employer effects (ψ). Also, there is very little variation in estimated employer effects between classes 3 and 8. As we will see, our model only detects four distinct employer classes. All of these posterior distributions are extremely tight around the posterior mean.

Table 1 reports the posterior mean and MCSE of the parameter governing the population distribution of worker types, π_A , the population distribution of employer type, π_B , and the marginal probability for match type, π_K . The latter probability is computed by integrating the conditional

Table 1: Posterior Distribution of Worker, Firm, and Match Population Heterogeneity

	(1) Worker		(2) Firm		(3) Match		
	Mean	MCSE	Mean	MCSE	Mean		
π_{A1}	0.0875	0.0078	π_{B1}	0.2038	0.2668	π_{K1}	0.0093
π_{A2}	0.1317	0.0127	π_{B2}	0.5985	0.2682	π_{K2}	0.0249
π_{A3}	0.1448	0.0032	π_{B3}	0.0003	0.0001	π_{K3}	0.0916
π_{A4}	0.1379	0.0094	π_{B4}	0.0003	0.0001	π_{K4}	0.0616
π_{A5}	0.0988	0.0103	π_{B5}	0.0003	0.0001	π_{K5}	0.0509
π_{A6}	0.1005	0.0064	π_{B6}	0.0002	0.0001	π_{K6}	0.0620
π_{A7}	0.0835	0.0052	π_{B7}	0.0002	0.0000	π_{K7}	0.1014
π_{A8}	0.0695	0.0034	π_{B8}	0.0159	0.0114	π_{K8}	0.1572
π_{A9}	0.0968	0.0067	π_{B9}	0.0203	0.0129	π_{K9}	0.3463
π_{A10}	0.0490	0.0017	π_{B10}	0.1602	0.0049	π_{K10}	0.0948

Columns (1) and (2) report the posterior distribution of the estimated probability that a worker (firm) belongs to each latent heterogeneity class. The classes are sorted in increasing order by the associated wage heterogeneity component. Columns (3) report the marginal distribution across the population of latent match heterogeneity classes. There are no associated MCSE estimates because the marginal distribution is computed from the conditional distribution of match quality given the worker and firm types.

probability, $\pi_{k|ab}$ over worker and employer types. All worker types occur in the population with positive probability, however, workers are more likely to be of the highest and lowest type in the population. The case is even more extreme for employers. The distribution of employer types has only three points of support with non-negligible mass. The distribution of employer types is thus very coarse. By contrast, the distribution of match classes is the most granular. The marginal distribution of match types is clearly skewed toward higher match quality.

For completeness, Table OA1 in Appendix E reports the posterior mean and MCSE for the parameters associated with observed covariates included in the earnings model. These estimates are qualitatively similar to the uncorrected estimates. Whatever correcting for endogenous mobility does to estimated worker and employer effects, it seems to have no substantive effect on estimates of time-varying observable characteristics.

Comparison of AKM and Structural Estimates

Table 2 reports correlations, weighted by job duration, among earnings and its components as estimated by least squares (labeled AKM) and from our structural endogenous mobility model (labeled Gibbs). Column (1) reports the correlation of earnings (labeled $\ln w$) with each of the earnings components. In the structural estimates, much more of the variation in earnings is explained by individual heterogeneity than in the AKM estimates. Much less of the variation is explained by employer and match specific heterogeneity. As we will see, though, this is partially due to a strong negative correlation between the structural match effect and the structural firm and worker effects.

Table 2: Correlation Matrix of Wage Equation Parameters: LEHD Data 0.5% Sample, (10, 10, 10) Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	$\ln w$	$X\beta_{AKM}$	θ_{AKM}	ψ_{AKM}	μ_{AKM}	ε_{AKM}	$X\beta_{AKM}$	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}
$\ln w$	1.00										
$X\beta_{AKM}$	0.44	1.00									
θ_{AKM}	0.39	-0.49	1.00								
ψ_{AKM}	0.50	0.07	0.17	1.00							
μ_{AKM}	0.34	0.03	0.00	-0.00	1.00						
ε_{AKM}	0.20	-0.02	0.00	0.00	-0.00	1.00					
$X\beta_{Gibbs}$	0.78	0.56	0.25	0.24	0.04	-0.02	1.00				
θ_{Gibbs}	0.50	0.14	0.38	0.27	0.00	0.00	0.25	1.00			
ψ_{Gibbs}	0.27	0.02	0.12	0.42	0.11	0.00	0.10	0.04	1.00		
μ_{Gibbs}	0.06	0.05	-0.05	-0.10	0.28	0.00	-0.00	-0.23	-0.74	1.00	
ε_{Gibbs}	0.27	0.00	0.02	0.08	0.17	0.78	0.00	0.00	0.00	0.00	1.00

Table entries are means of the correlation between the indicated variables across 7,922 draws from the Gibbs sampler described in the text.

The AKM estimates of the person and firm effects have a positive correlation of 0.1665. This result contrasts somewhat with prior estimates from LEHD data that reported a correlation much closer to zero (Abowd et al. 2003; 2012). This discrepancy may reflect changes over time in the nature of assortative matching, as seems to have been the case in Germany Card et al. (2013). We also have a much longer sample than was available in the prior studies (1990–2012 versus 1990–2003), which attenuates any limited mobility bias Andrews et al. (2012). The AKM match effect, because it is estimated from the least squares residual is restricted to be uncorrelated with the AKM 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 structural match effect.

In the structural estimation, the correlation between person and firm effects is much weaker, at 0.0430. Note that this need not be the case: our structural model allows for separation and assignment outcomes to be arbitrarily associated with worker and firm types. In the structural model, there is a very strong negative correlation between the structural match effect and the structural worker effect (-0.2271) and the structural firm effect (-0.7350). The structural estimates also exhibit a positive correlation between observed and unobserved components of individual earnings heterogeneity: $\text{Corr}(X\beta_{Gibbs}, \theta_{Gibbs}) = 0.2495$. This result contrasts with the AKM estimates, for which $\text{Corr}(X\beta_{AKM}, \theta_{AKM}) = -0.4865$.

The bottom-left panel of Table 2 reports the correlation between the inconsistent AKM parameters and the structural parameters. This panel provides some insight into the endogenous mobility bias. Observe, first, there is a positive correlation between the AKM heterogeneity components and their structural counterparts. Second, the AKM estimates of the worker and firm effects combine information from the structural worker, firm, and match effects. In particular, the AKM person effect is negatively correlated with the structural match effect, as is the AKM firm effect.

Table OA2 in Appendix E reports the same information in the form of a 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 used to compute endogenous mobility-corrected estimates of the wage components from data for which only the AKM estimates are available.

These results indicate that the positive correlation between the AKM person and firm effects do not actually reflect positive assortative matching on latent worker and employer characteristics. They arise instead because high wage workers are typically employed on low paying matches, and the heterogeneity from those matches loads onto the OLS worker and firm effects. The structural model indicates that workers are assigned randomly to firms, and that most earnings variation is associated with worker and match-specific heterogeneity. However, that match-specific heterogeneity acts to compress the wage distribution.

5.5 Summary of Structural Mobility Model

Figures 5 and 6 summarize the mobility model by presenting the stationary distribution of worker-firm pairs and the expected match effect on each such pair. We obtain the steady-state distribution by computing the kernel of the Markov transition matrix implied by our mobility model from the estimated parameters, γ , δ , π_A , π_B and $\pi_{K|AB}$. This gives us, for each worker type, the steady-state probability of observing a worker of that type matched to a particular type firm on a particular type of match.

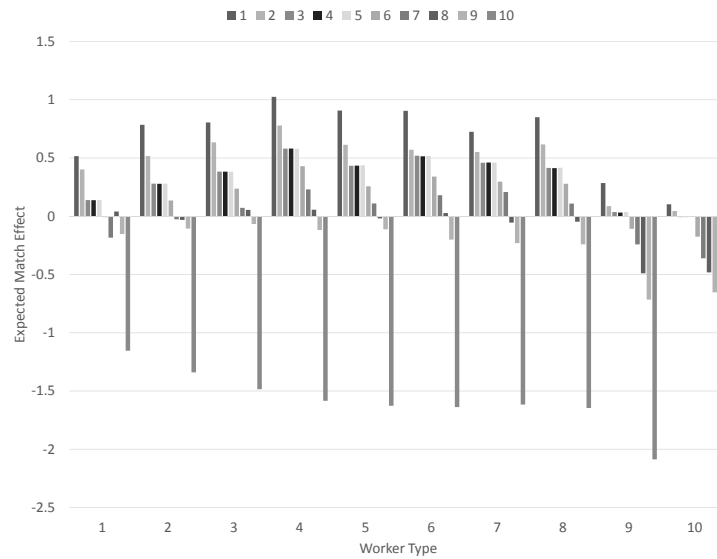


Figure 5: Expected structural match effect in steady-state by worker/firm type cells. Legend is firm types.

Figure 5 is structured identically to its AKM analog (Figure 2). Each bar represents the expected match effect, conditional on a worker type–firm type cell. The bars are grouped by worker type, so that within each worker type, we see the pattern of expected match effects for workers of that type when matched to different firm types. The figure shows two key patterns: the expected match effect is strongly decreasing with firm type, and weakly decreasing with worker type. Thus, the mobility model in steady-state exhibits the same compression pattern we observed in Table 2, as anticipated in our discussion of the AKM estimates.

Figure 5 tells us what the expected match effect will be conditional on observing a given

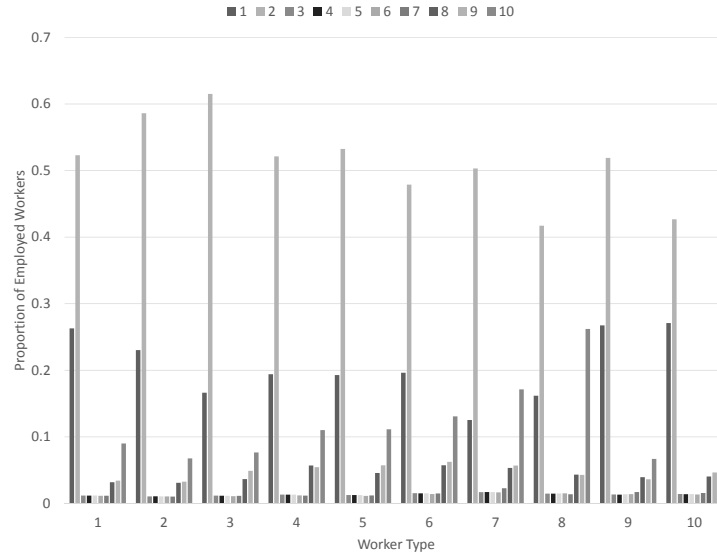


Figure 6: Expected share of matches in steady-state for each worker-type firm-type combination. Legend is firm types.

worker-firm combination. Figure 6 gives the probability of observing each worker-firm combination in steady-state. Again, the data are grouped by worker type, so that within each group, the bars show the probability of observing a worker in a firm of each type. The pattern for each worker type is very similar to the population distribution of firm types reported in Table 1. There is some evidence of a selection effect: workers of type 10 are roughly 50 percent more likely to be observed in high-wage firms as workers of type 1. This is consistent with the weak evidence of positive assortative matching from Table 2. Overall, it appears as though worker-firm matches are sampled almost randomly in steady-state.

Figure 7 shows the probability a worker is observed in non-employment, conditional on the latent worker type. For comparison, we also show the corresponding probabilities when workers are grouped by the decile of their AKM worker effect. The estimates from the structural model show the employment probability is increasing in the latent worker type. That is, more highly paid workers are more likely to be employed. This standard mover-stayer result is not assumed in estimation. By contrast, the results based on the AKM worker effect deciles indicate that highly-paid and low-paid workers are less likely to be employed than workers in the middle of the distribution. The non-employment probabilities may seem high because they incorporate transitions out of the

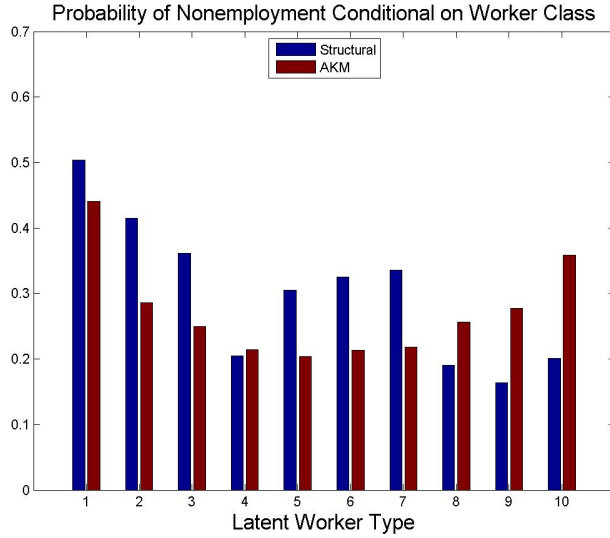


Figure 7: Probability of non-employment by latent worker type (Gibbs) and worker effect decile (AKM)

labor force and into uncovered employment in addition to unemployment.

Finally, Figure OA1 in Appendix E reports the structural analog to Figure 3. It reports, for workers who leave their job, the expected match effect on the origin job conditional on the worker type, and on the type of the origin and destination job. Recall there are very few firms of types 3–7. The destination firm type is not independent of the match effect on the origin job, but the predictive power of the match effect actually seems much weaker than was the case in Figure 3 based on the inconsistent AKM estimates.

5.6 Validation: Relationship with Revenue

If the AKM estimates are biased by endogenous mobility, we should observe a difference in the relationship between employer revenues and the earnings heterogeneity components estimated from our structural model. Table 3 reports the results of estimating a firm-level regression of log revenue per worker onto the estimated firm effect, the average worker effect, and the average match effect. For this analysis, we restrict the sample to jobs from 2002, which is the year the revenue data were reported in the Economic Census. After this restriction, the sample consists of 60,116 firms. The columns under (1) report the parameter estimates and standard errors from the regression onto firm-

level aggregates of the structural wage components. The columns under (2) reports the estimates from the regression onto firm-level aggregates of the uncorrected AKM wage components.

Table 3: Regression of Log Revenue Per Worker on Structural and AKM Estimates of Wage Decomposition Components

	(1) Gibbs		(2) AKM	
	Coef	Ste	Coef	Ste
Firm Avg. θ	0.2288	0.0119	0.0234	0.0059
ψ	0.2431	0.0082	0.6735	0.0133
Firm Avg. μ	0.2046	0.009	0.0158	0.0094
Firm Avg. $X\beta$ or Exper	0.0343	0.0063	-0.0231	0.0045
Intercept	3.4898	0.0532	3.9476	0.0247
N	60, 116		60, 116	

Coefficient estimates from firm-level regressions of log revenue per worker onto firm-level averages of estimated earnings heterogeneity. Column (1) uses the estimated wage components corrected for endogenous mobility bias from the structural model. Column (2) uses the uncorrected estimates computed using the AKM decomposition under the assumption of exogenous mobility. The sample is based on data from 2002, the year for which revenue data are reported in the 2002 Economic Census.

When we use the AKM components of earnings heterogeneity, revenue per worker is strongly correlated with the AKM firm effect, and only negligibly correlated with the average worker and match effects. When, by contrast, we use the estimates of earnings heterogeneity from the endogenous mobility model, the average match effect and average worker effect are much more strongly correlated with revenue. We also find that the correlation between the structural firm effect and log revenue per worker is considerably smaller than when the AKM firm effect is used: 0.2431 versus 0.6735.

We interpret these results as supporting our correction for endogenous mobility. Revenue per worker should be a function of total human capital, both in the form of average worker quality and match quality. In the uncorrected estimates, there is no relationship between worker quality, match quality, and revenue. The variation in revenue per worker, which drives workers across jobs, all loads onto the firm effect. After the correction for endogenous mobility, worker heterogeneity and match quality have the relationships we would expect with revenue per worker. It remains the case that more productive firms – those with greater revenue per worker – are also high wage firms.

6 Conclusion

Our analysis confirms the relevance of relaxing the assumption that job mobility is exogenous. Exogenous mobility is entailed by the AKM earnings decomposition, as well as by all downstream applications that use the estimated worker and firm effects as measures of skill and compensation policy. We have shown that the assumption of exogenous mobility is rejected in data from the LEHD program. Furthermore, the analysis of residuals from the AKM decomposition indicate the presence of omitted match-specific heterogeneity, and that omitted match-specific heterogeneity is predictive of the type of firm to which a worker moves.

To relax the exogenous mobility assumption, we estimate a latent class model that incorporates these features. We allow for a match effect that is arbitrarily correlated with worker and firm heterogeneity, and we allow the match effect to drive both the decision to separate and the type of firm to which the worker moves. Our results indicate that allowing for correlated match effects has a strong effect on estimated worker and firm-specific heterogeneity. Validation against firm revenue data suggest that our corrections for endogeneity move the estimated effects in an economically reasonable direction.

Our analysis is subject to some caveats. Our model is extremely computationally intensive, requiring us to estimate on a sample of the LEHD data. While we have used a dense subsample, the implications of sampling in relational, or network data, particularly for this sort of analysis remain poorly understood. The parallelization afforded by the graph coloring algorithm can be scaled up, but not within the computing facilities available through the Census Bureau. Additionally, we have considered one model that relaxes the exogenous mobility assumption. The model we consider is consistent with the residual diagnostics, but other models are possible.

For researchers working with these models, our results indicate that it is important to test for failure of the exogenous mobility assumption. When it fails, it may also be important to attempt to correct for endogenous mobility bias. We find that the OLS estimates from the AKM model are positively correlated with conceptually appropriate effects that were estimated correcting for endogenous mobility. However, correcting for endogenous mobility has a substantial affect on the relationship between worker and firm earnings heterogeneity, and on the relationship of these components with firm revenue per worker. These findings recommend that caution is warranted

when interpreting worker and firm effects estimated under the AKM assumptions, as originally noted by Abowd and Kramarz (1999).

Acknowledgements

We have benefitted from discussion with Joseph Altonji, David Card, Patrick Kline, Francis Kramarz, Jesse Rothstein, Richard Upward, and Lars Vilhuber, along with participants at numerous seminars. Nellie Zhao provided expert research assistance. The research in this paper was previously circulated in two working papers: “Endogenous Mobility”, and “How Important is Endogenous Mobility for Measuring Employer and Employee Heterogeneity”.

Abowd acknowledges direct support from NSF Grants SES-0339191, CNS-0627680, SES-0922005, TC-1012593, and SES-1131848. Some of the research for this paper was conducted using the resources of the [Social Science Gateway](#), which is partially supported by NSF grant SES-0922005. This paper was written while the first author was Distinguished Senior Research Fellow and the third 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. An archive of the code used to prepare analysis samples and conduct estimation along with the released results is available directly from <http://digitalcommons.ilr.cornell.edu/ldi/23>.

Supplementary Materials

AMS_EndMob_RepArchive_Tests_v2015May.zip: Archive of SAS code to perform the tests described in the article. The archive also includes the output files released by the Census Bureau after disclosure avoidance review. (zipped archive)

AMS_EndMob_RepArchive_Model_v2015May.zip: Archive of SAS and MATLAB code to build the analysis data, estimate the models, and generate the model summaries described in the article. The archive also includes the files summarizing the model output released by the Census Bureau after disclosure avoidance review. (zipped archive)

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. and Kramarz, F. (1999). The analysis of labor markets using matched employer-employee data, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3 of *Handbook of Labor Economics*, Elsevier, chapter 40, pp. 2629–2710.

Abowd, J. M., Kramarz, F., Lengerman, P., McKinney, K. and Roux, S. (2012). Persistent inter-industry wage differences: Rent-sharing and opportunity costs, *IZA Journal of Labor Economics* **1**(1): 1–7.

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., 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). The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators, in T. Dunne, J. Jensen and M. Roberts (eds), *Producer Dynamics: New Evidence from Micro Data*, Chicago: University of Chicago Press for the National Bureau of Economic Research, pp. 149–230.

- Andrews, M. J., Gill, L., Schank, T. and Upward, R. (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias, *IZA Discussion Papers 6662*, Institute for the Study of Labor (IZA).
- Card, D., Heining, J. and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality, *The Quarterly Journal of Economics* **128**(3): 967–1015.
- 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.
- Gebremedhin, A. H., Manne, F. and Pothen, A. (2005). What color is your jacobian? Graph coloring for computing derivatives, *SIAM Rev.* **47**(4): 629–705.
- Geyer, C. (2011). Introduction to Markov Chain Monte Carlo, *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, Chapman and Hall/CRC, pp. –.
- Geyer, C. J. (1992). Practical Markov Chain Monte Carlo, *Statistical Science* **7**(4): pp. 473–483.
- Gibbons, R., Katz, L. F., Lemieux, T. and Parent, D. (2005). Comparative advantage, learning and sectoral wage determination, *Journal of Labor Economics* **23**: 681–723.
- Helwege, J. (1992). Sectoral shifts and interindustry wage differentials, *Journal of Labor Economics* **10**(1): 55–84.
- 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.

- Kosorok, M. R. (2000). Monte carlo error estimation for multivariate markov chains, *Statistics & Probability Letters* **46**(1): 85 – 93.
- Kramarz, F., Machin, S. and Ouazad, A. (2014). Using compulsory mobility to identify school quality and peer effects, *Oxford Bulletin of Economics and Statistics* .
- Krishna, P., Poole, J. P. and Senses, M. Z. (2014). Wage effects of trade reform with endogenous worker mobility, *Journal of International Economics* **93**(2): 239 – 252.
- 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. M. (2014). Free to move? A network analytic approach for learning the limits to job mobility, *Labour Economics* **29**: 49 – 61.
- Schmutte, I. M. (2015). Job referral networks and the determination of earnings in local labor markets, *Journal of Labor Economics* **33**(1): 1–32.
- 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.
- Woodcock, S. (2008). Wage differentials in the presence of unobserved worker, firm, and match heterogeneity, *Labour Economics* **15**: 771–793.

Appendices

A Online Appendix: Formal Test of Endogenous Mobility

We implement our tests using LEHD data covering 28 states over the period 1990–2004. Our estimation sample for the AKM decomposition consists of all 982 million dominant job observations in this time period. Recall from the main text that the dominant job is the one from which the individual earns the most labor market earnings in a particular year. To implement the tests, we discretize estimated person effects, firm effects, and residuals onto a fixed support. The quantiles that define the support points are calculated from a point-in-time snapshot of the distribution of dominant jobs in progress as of April 1, 2002. That distribution is restricted to full-time, full-year jobs held by individuals age 18-70. Finally, in testing, we use all 465 million dominant job observations for workers 18-70 that occur between 1999 and 2002. Test 1, the match effects test, uses data for about 104 million job changers during 1999-2004, inclusive. Test 2, the productive workforce test, uses data for about 4 million firms alive in 2001.

A.1 Data Preparation and Definitions

Given the fitted values from the AKM decomposition, we select the sample of individuals and employers active at the beginning of 2002, quarter 2 (April 1, 2002). For this sample, we compute deciles from the estimated $\hat{\theta}_i$, $\hat{\psi}_{J(i,t)}$, and $\hat{\varepsilon}_{it}$ as described above. Using the estimated deciles, we discretize each component of the decomposition onto 10 fixed points of support. We adopt the following notation:

$$Q(z) = a \text{ denotes quantile } a \text{ for } z \in \{\theta, \psi, \varepsilon\}$$

and

$$\#Q(z) \text{ denotes then number of quantiles for } z \in \{\theta, \psi, \varepsilon\}.$$

In the tests presented below, we use deciles, so $\#Q(z) = 10$.

A.2 Test Statistic 1: Match Effects Test

Under the hypothesis of exogenous mobility, the match effect for a given individual–employer pair can be estimated using the average residual for the most recent completed job at j by i . We denote these match effects as $\bar{\varepsilon}_{it-1}$ for those individuals who change employers between periods $t - 1$ and t . Formally,

$$\bar{\varepsilon}_{it-1} = \frac{\sum_{\{s|J(i,s)=j \wedge s < t \wedge J(i,s) \neq J(i,t)\}} \hat{\varepsilon}_{is}}{\sum 1_{\{s|J(i,s) = j \wedge s < t \wedge J(i,s) \neq J(i,t)\}}}$$

An individual for whom $\bar{\varepsilon}_{it-1} > 0$ received wage payments while employed at $J(i, t - 1) = j$ that exceeded their expected value, again under the hypothesis of exogenous mobility. The opposite is true for individuals for whom $\bar{\varepsilon}_{it-1} < 0$.

A.2.1 Derivation of the Match Effects Test Statistic

To form a test statistic that captures the potential for $\bar{\varepsilon}_{it-1}$ to be predictive of the next employer type, we count all (i, t) pairs where $J(i, t-1) \neq J(i, t)$ (job changers) in quantiles of the components $\hat{\theta}_i$, $\hat{\psi}_{J(i,t-1)}$, $\hat{\psi}_{J(i,t)}$, and $\bar{\varepsilon}_{it-1}$:

$$n_{abcd} = \sum_{\{i,t|J(i,t-1) \neq J(i,t)\}} 1 \left\{ \begin{array}{l} Q(\hat{\theta}_i) = a \wedge \\ Q(\hat{\psi}_{J(i,t-1)}) = b \wedge \\ Q(\hat{\psi}_{J(i,t)}) = c \wedge \\ Q(\bar{\varepsilon}_{it-1}) = d \end{array} \right\}. \quad (\text{A-1})$$

The joint probability of observing n_{abcd} is

$$\pi_{abcd} = \Pr \{ Q(\theta_i) = a \wedge Q(\psi_{J(i,t-1)}) = b \wedge Q(\psi_{J(i,t)}) = c \wedge Q(\bar{\varepsilon}_{it-1}) = d \}.$$

Exogenous mobility implies that the match effect from period $t-1$ should not be predictive of the transition from $\psi_{J(i,t-1)}$ to $\psi_{J(i,t)}$ for an individual with θ_i . This hypothesis can be formalized as conditional independence of the outcome

$$(Q(\theta_i) = a \wedge Q(\psi_{J(i,t-1)}) = b \wedge Q(\psi_{J(i,t)}) = c)$$

from $Q(\bar{\varepsilon}_{it-1}) = d$. In terms of the joint probabilities we compute

$$X_{\nu_1}^2 = \text{Test}(\pi_{abcd} = \pi_{abc+} \pi_{+++d}) \quad (\text{A-2})$$

where the subscript $+$ denotes the marginal distribution with respect to the indicated dimension, and degrees of freedom are given by

$$\nu_1 = (\#(Q(\theta_i)) \times \#Q(\psi_{J(i,t-1)}) \times \#Q(\psi_{J(i,t)}) - 1) \times (\#Q(\bar{\varepsilon}_{it-1}) - 1).$$

A.2.2 Computation of the Match Effects Test

We compute the test statistic (A-2) by direct calculation of the chi-squared statistic from the 4-way contingency table defined by the discretized earnings heterogeneity under the conditional independence assumption $\pi_{abcd} = \pi_{abc+} \pi_{+++d}$. The population of job changers consists of individuals i for whom $J(i, t-1) \neq J(i, t)$ for $t = 1999, \dots, 2003$. The entire population of individuals and employers was used to compute the quantiles of the $\hat{\theta}_i$, $\hat{\psi}_{J(i,t-1)}$, $\hat{\psi}_{J(i,t)}$, and $\bar{\varepsilon}_{it-1}$ distributions. Then the counts (A-1) were tabulated using all observations in the job-changer population and used to compute the relevant marginal frequencies for the test.

A.3 Test Statistic 2: Productive Workforce Test

Our second test considers the implications of exogenous mobility for the employer's choice of workforce distributions over θ_i . The average amount by which wages deviate from their expectations, under exogenous mobility, for a given workforce at a point in time can be computed as the average residual for all employees at $J(i, t) = j$ in year t

$$\tilde{\varepsilon}_{jt} = \frac{\sum_{\{i|J(i,t)=j\}} \hat{\varepsilon}_{it}}{\sum 1_{\{i|J(i,t)=j\}}}.$$

An employer for whom $\tilde{\varepsilon}_{jt} > 0$ has paid higher than expected wages in period t ; and the opposite is true for $\tilde{\varepsilon}_{jt} < 0$. Although there could be many reasons for this, we will refer to $\tilde{\varepsilon}_{jt}$ as a measure of workforce productivity. However, the exogenous mobility hypothesis is silent about the meaning of $\tilde{\varepsilon}_{jt}$. What matters is its relationship to the within-employer distribution of θ_i . If $\tilde{\varepsilon}_{js}$ is predictive of the within-employer distribution of θ_i for some period $t > s$, given ψ_j , then exogenous mobility fails because the distribution of future employment depends on residuals in the theoretical AKM decomposition.

To implement this test, consider two periods $s < t$ and all employers with strictly positive employment in period s . Compute the counts

$$n_{abc|s} = \sum_j \left\{ 1_{\{Q(\psi_j) = a \wedge Q(\tilde{\varepsilon}_{js}) = c\}} \times \sum_{\{i|J(i,s)=j \wedge Q(\psi_j)=a\}} Q(\theta_i) = b \right\}$$

and

$$n_{abc|t} = \sum_j \left\{ 1_{\{Q(\psi_j) = a \wedge Q(\tilde{\varepsilon}_{js}) = c\}} \times \sum_{\{i|J(i,t)=j \wedge Q(\psi_j)=a\}} Q(\theta_i) = b \right\}.$$

Note that the two counts are not independent because they condition on the same distribution of employers alive in period s . Let

$$\pi_{abc|s} = \Pr \{Q(\psi_j) = a \wedge (Q(\theta_i) = b|s) \wedge Q(\tilde{\varepsilon}_{js}) = c\}$$

and

$$\pi_{abc|t} = \Pr \{Q(\psi_j) = a \wedge (Q(\theta_i) = b|t) \wedge Q(\tilde{\varepsilon}_{js}) = c\}.$$

Then, the statistic for testing the conditional independence of the within-employer distribution over θ_i with respect to the residual is

$$X_{\nu_2}^2 = Test \left(\ln \left(\frac{\pi_{abc|s}}{\pi_{abc|t}} \right) = \ln \left(\frac{\pi_{ab+|s}}{\pi_{ab+|t}} \right) \right) \quad (\text{A-3})$$

with degrees of freedom $\nu_2 = (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1) + (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$.

A.3.1 Derivation of the Productive Workforce Test Statistic

To see why the test in equation (A-3) is correct, consider the log-linear model

$$\begin{aligned} \ln \left(\frac{\pi_{abc|s}}{\pi_{abc|t}} \right) &= (\mu_{a|s} - \mu_{a|t}) + (\mu_{b|s} - \mu_{b|t}) + (\mu_{c|s} - \mu_{c|t}) \\ &\quad + (\gamma_{ab|s} - \gamma_{ab|t}) + (\gamma_{ac|s} - \gamma_{ac|t}) + (\gamma_{bc|s} - \gamma_{bc|t}) \\ &\quad + (\rho_{abc|s} - \rho_{abc|t}) \end{aligned}$$

where the notation is as follows:

- $\mu_{z|t}$ denotes main effects of $z \in \{Q(\psi_j), Q(\theta_i), Q(\tilde{\varepsilon}_{js})\}$ in period t ,
- $\gamma_{yz|t}$ denotes 2-way interactions of $(y, z) \in \{Q(\psi_j), Q(\theta_i), Q(\tilde{\varepsilon}_{js})\}$ in period t ,
- $\rho_{xyz|t}$ denotes 3-way interactions of $(x, y, z) \in \{Q(\psi_j), Q(\theta_i), Q(\tilde{\varepsilon}_{js})\}$ in period t .

The change in main effects of $Q(\psi_j)$ from period s to t , $(\mu_{a|s} - \mu_{a|t})$, must be 0 since the population of employers is restricted to be identical in both periods. Similarly, the change in main effects of $Q(\tilde{\varepsilon}_{js})$, $(\mu_{c|s} - \mu_{c|t})$, must be 0 since the workforce productivity distribution is only measured at period s . The change in interaction of $Q(\psi_j)$ and $Q(\tilde{\varepsilon}_{js})$, $(\gamma_{ac|s} - \gamma_{ac|t})$, must also be 0 for the same reason.

This leaves two sets of parameters that are unconstrained by the null hypothesis—the change in main effects of $Q(\theta_i)$, $(\mu_{b|s} - \mu_{b|t})$, with $df = (\#Q(\theta_i) - 1)$ and the change in interaction of $Q(\psi_j)$ and $Q(\theta_i)$, $(\gamma_{ab|s} - \gamma_{ab|t})$, with $df = (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1)$. The parameters affected by the null hypothesis are the change in interaction of $Q(\theta_i)$ and $Q(\tilde{\varepsilon}_{js})$, $(\gamma_{bc|s} - \gamma_{bc|t})$, with $df = (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$ and the change in interaction of $Q(\psi_j)$, $Q(\theta_i)$ and $Q(\tilde{\varepsilon}_{js})$, $(\rho_{abc|s} - \rho_{abc|t})$, with $df = (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$. Under the null hypothesis $(\gamma_{bc|s} - \gamma_{bc|t}) = 0$ and $(\rho_{abc|s} - \rho_{abc|t}) = 0$ with $df = \nu_2 = (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1) + (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$.

A.3.2 Computation of the Productive Workforce Test Statistics

We use the method of moments for test (A-3). The observations are firms j with positive employment in s . For each firm compute

$$x_j = \begin{bmatrix} \frac{n_{j1t}}{n_{j+t}} - \frac{n_{j1s}}{n_{j+s}} \\ \frac{n_{j2t}}{n_{j+t}} - \frac{n_{j2s}}{n_{j+s}} \\ \dots \\ \frac{n_{j(\#Q(\theta_i)-1)t}}{n_{j+t}} - \frac{n_{j(\#Q(\theta_i)-1)s}}{n_{j+s}} \end{bmatrix}$$

where

$$n_{jqt} = \sum_{\{i|J(i,t)=j\}} 1(Q(\theta_i) = q).$$

and x_j is $[(\#Q(\theta_i) - 1) \times 1]$. For each value of a and c compute the vector of means and the covariance matrix

$$\bar{x}_{ac} = \frac{\sum_{\{j|Q(\psi_j)=a \wedge Q(\tilde{\epsilon}_{js})=c\}} n_{j+s} x_j}{\sum_{\{j|Q(\psi_j)=a \wedge Q(\tilde{\epsilon}_{js})=c\}} n_{j+s}}$$

$$V_{ac} = \frac{\sum_{\{j|Q(\psi_j)=a \wedge Q(\tilde{\epsilon}_{js})=c\}} n_{j+s} (x_j - \bar{x}_{ac})(x_j - \bar{x}_{ac})'}{\sum_{\{j|Q(\psi_j)=a \wedge Q(\tilde{\epsilon}_{js})=c\}} n_{j+s}}.$$

$$N = \sum_j 1(j|\exists i : J(i, s) = j)$$

For each value of a compute the expected mean under the null hypothesis

$$\bar{x}_a = \frac{\sum_{\{j|Q(\psi_j)=a\}} n_{j+s} x_j}{\sum_{\{j|Q(\psi_j)=a\}} n_{j+s}}.$$

Then,

$$X_{\nu_2}^2 = N \sum_{a,c} (\bar{x}_{ac} - \bar{x}_a)' V_{ac}^{-1} (\bar{x}_{ac} - \bar{x}_a).$$

Under the null hypothesis, $X_{\nu_2}^2 = 0$ and follows a chi-square distribution with ν_2 degrees of freedom.

B Online Appendix: Posterior Distribution of the Parameter Vector

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

$$p(\rho|Y, Z) \propto \mathcal{L}(\rho|Y, Z) \frac{1}{\sigma^{\nu_0+1}} \exp\left(-\frac{s_0^2}{\sigma^2}\right) \prod_{\ell=1}^L \pi_{a\ell}^{\frac{1}{L}-1} \prod_{m=1}^M \pi_{bm}^{\frac{1}{M}-1} \quad (\text{B-1})$$

$$\times \prod_{\ell=1}^L \prod_{m=0}^M \prod_{q=1}^Q \left(\pi_{q|\ell m}^{\frac{1}{Q}-1} \gamma_{\ell m q}^{\frac{1}{2}-1} (1 - \gamma_{\ell m q})^{\frac{1}{2}-1} \prod_{m'=0}^M \delta_{m'|\ell m q}^{\frac{1}{M+1}-1} \right).$$

This distribution factors into posterior distributions for the model parameters that are independent, conditional on the latent data, from which we sample.

To characterize these distributions, we introduce new notation. The matrix $G = [X \ A \ B \ K]$ is the full design of observed characteristics, ability, productivity, and match types given the observed and latent data. The term ν , which appears in the posterior of σ , is $\nu = N + \nu_0 - (L + M + Q)$. The sum of squared log earnings residuals is

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

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

Key to estimation are 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, $n_{m'|\ell m q}^{trans}$ 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' .

The posterior distribution of the wage equation parameters is

$$\begin{bmatrix} \alpha \\ \beta \\ \theta \\ \psi \\ \mu \end{bmatrix} | \sigma \sim N \left(\begin{bmatrix} \hat{\alpha} \\ \beta \\ \theta \\ \psi \\ \mu \end{bmatrix}, \sigma^2 (G^T G)^{-1} \right) \quad (\text{B-3})$$

where

$$\begin{bmatrix} \hat{\alpha} \\ \beta \\ \theta \\ \psi \\ \mu \end{bmatrix} = (G^T G)^{-1} G^T w,$$

and

$$\sigma^2 \sim \text{IG} \left(\frac{\nu}{2}, \frac{2}{\nu S^2} \right). \quad (\text{B-4})$$

The posterior distributions for the latent heterogeneity types are Dirichlet:

$$\pi_a \sim \text{D} \left(n_{a1} + \left(\frac{1}{L} - 1 \right), \dots, n_{aL} + \left(\frac{1}{L} - 1 \right) \right); \quad (\text{B-5})$$

$$\pi_b \sim \text{D} \left(n_{b1} + \left(\frac{1}{M} - 1 \right), \dots, n_{bM} + \left(\frac{1}{M} - 1 \right) \right); \quad (\text{B-6})$$

$$\pi_{k|ab} \sim \text{D} \left(n_{k|ab1} + \left(\frac{1}{Q} - 1 \right), \dots, n_{k|abQ} + \left(\frac{1}{Q} - 1 \right) \right). \quad (\text{B-7})$$

The posterior distributions of the separation and assignment parameters of the mobility model are also Dirichlet:

$$\gamma_{lmq} \sim \text{D} \left(n_{lmq}^{sep} + \left(\frac{1}{2} - 1 \right), n_{lmq}^{stay} + \left(\frac{1}{2} - 1 \right) \right); \quad (\text{B-8})$$

$$\delta_{b|lmq} \sim \text{D} \left(n_{0|lmq}^{trans} + \left(\frac{1}{M+1} - 1 \right), \dots, n_{M|lmq}^{trans} + \left(\frac{1}{M+1} - 1 \right) \right). \quad (\text{B-9})$$

C Online Appendix: The Mobility Model in Steady-State

The stationary distribution of the mobility model gives a steady-state distribution of employment spells across worker, employer, and match types. This, it turns out, is a model for the realized mobility network, characterized in the data by the design matrix of employer effects, F , and the associated cross-product term, $D^T F$. We also interpret it as a characterization of the selection model – the process by which particular matches are selected from the set of all possible matches.

The stationary distribution is simple to characterize: define $\lambda_{\ell,m,q}$ to be the expected number of matches in steady-state between workers of type ℓ and employers of type m on matches with quality q . Now define the diagonal matrix

$$\Lambda = \text{diag}([\lambda_{111}, \lambda_{112}, \dots, \lambda_{LMQ}]^T). \quad (\text{C-1})$$

Note that Λ does not account for transitions to non-employment. For exposition, suppose $L = M = Q = 2$ so Λ is 8×8 . In estimation, we let L , M , and Q vary and report results for the case

$L = Q = M = 10$.

In steady-state, observed log earnings data $\ln w$ are drawn from a discrete distribution proportional to Λ . Net of the statistical residual, and the effect of observed time-varying characteristics, $X\beta$, the potential outcomes $\ln w - X\beta - \varepsilon$ are completely characterized by an $LMQ \times 1$ vector, \tilde{y} with

$$\tilde{y}_{\ell,m,q} = \alpha + \theta_\ell + \psi_m + \mu_q. \quad (\text{C-2})$$

The model therefore specifies

- Potential Outcomes: \tilde{y} , and
- Selection Process: Λ .

Define a set of indicator matrices analogous to the person, employer, and match design matrices. For the $2 \times 2 \times 2$ model, this matrix is simply

$$[\tilde{D} \quad \tilde{F} \quad \tilde{G}] = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}. \quad (\text{C-3})$$

The notation \tilde{D} , \tilde{F} , and \tilde{G} highlights the connection between these reduced-dimension objects and the design matrices of worker and employer effects in the full data AKM model.

Net of $X\beta$ and ε , the earnings data are sampled from a distribution proportional to

$$\Lambda \tilde{y} = \Lambda \left(\tilde{D}\theta + \tilde{F}\psi + \tilde{G}\mu \right) \quad (\text{C-4})$$

and the full cross-product matrix is

$$[\tilde{D} \quad \tilde{F} \quad \tilde{G}]^T \Lambda [\tilde{D} \quad \tilde{F} \quad \tilde{G}] = \begin{bmatrix} \tilde{D}^T \Lambda \tilde{D} & \tilde{D}^T \Lambda \tilde{F} & \tilde{D}^T \Lambda \tilde{G} \\ \tilde{F}^T \Lambda \tilde{D} & \tilde{F}^T \Lambda \tilde{F} & \tilde{F}^T \Lambda \tilde{G} \\ \tilde{G}^T \Lambda \tilde{D} & \tilde{G}^T \Lambda \tilde{F} & \tilde{G}^T \Lambda \tilde{G} \end{bmatrix}. \quad (\text{C-5})$$

Notice that the upper-left block of the cross-product matrix in (C-5) is a model for the Laplacian of the realized mobility network, which is random noise around this steady-state distribution.

D Online Appendix: Estimation Details

D.1 Parallelization of Employer Updates through Graph Coloring

To speed computation of the employer updates, we exploit the conditional independence restriction in the update formula, equation (15). For any employers, j and j' , we say j and j' are degree-one connected if any worker was observed to move from j directly to j' in the sample. The set, $N(j)$, is the set of all employers, j' , that are degree-one connected to j . Equation (15) implies that if j'' is not in $N(j)$ and j is not in $N(j'')$, then $\Pr[b_j = m | a, b_{-j}, k, Y, \rho]$ is independent of $\Pr[b_{j''} = m | a, b_{-j''}, k, Y, \rho]$ and, therefore, conditional on the rest of the latent data, the latent type of j and j'' can be updated at the same time (in parallel).

To fully exploit the network structure and conditional independence assumptions, we need groups of employers such that no two employers are degree-one connected. In the language of graph theory, this problem is equivalent to graph coloring in which the task is to color each node of a graph so that no two degree-one connected nodes have the same color, and to do so using the fewest colors possible.

For a general graph, the problem of finding the minimum number of colors is intractable. For our task, it is sufficient to find a coloring that yields a small number of partitions relative to the highest degree node in the data (well over 1,000). To that end, we implement the greedy sequential coloring algorithm described in Gebremedhin et al. (2005). Briefly, the algorithm sorts network nodes from highest to lowest degree (that is, sorting employers in descending order by the number of job-to-job separations). The first node is assigned a color at random. For every other node, we assign the least frequent color that has not already been applied to one of its neighbors. If there is no such color, we add a new color to the list and continue.

In our data, this algorithm yields a coloring that partitions employers into 24 non-intersecting subsets. We update the employer classes in parallel within each subset, and in sequence across the subsets. Our partition is well below the algorithmic worst-case guarantee: a coloring with as many colors as the highest-degree node in the graph, which is much greater than 1,000.

D.2 Calculation of Monte Carlo Standard Errors

When reporting results, we report Monte Carlo standard errors (MCSE) in place of, or in addition to, the posterior standard deviation. Unsurprisingly, we observe substantial autocorrelation across draws from the Gibbs sampler. The MCSE are computed using time-series methods that account for uncertainty about the location of the posterior distribution associated with autocorrelation in the chain. Using MCSE provides a practical and rigorous method for combining information across independent runs of the Gibbs sampler (we use three). The MCSE also fully exploit the information within each sample, while addressing within-thread autocorrelation, relative to more conventional ad hoc approaches like thinning the sample. Our ability to do so is all the more important given the computational burden of each draw. Even with the parallelization described in Section D.1, drawing from the Gibbs sampler is very time-consuming. Here, we describe implementation choices that affect our analysis. We refer the reader interested in the theoretical and practical details of computing the MCSE to the survey by Geyer (2011).

In calculating the MCSE, we implement the multivariate extension developed in Kosorok (2000) of initial sequence methods originally proposed by Geyer (1992). There are three variants of the initial sequence method, all of which exploit reversibility of the Markov Chain to determine the largest lag to include when computing the autocorrelation coefficient. The values reported in our tables are estimates from the initial positive sequence method, which are the most conservative. The other two methods, which we also implement, are the initial monotone and initial convex sequence methods. There is no meaningful difference across the estimates. In practice, we compute the univariate MCSE for each parameter due to numerical instability in the auto-covariance matrices.

E Online Appendix: Supplemental Tables and Figures

Table OA1: Posterior Mean of Observed Covariate Parameters

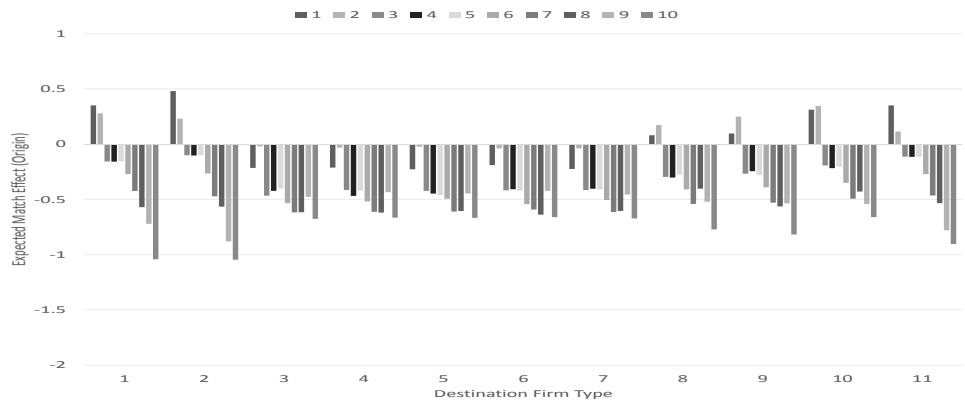
	Variable	Mean	(MCSE)	Variable	Mean	(MCSE)
	<i>age</i>	0.5810	(.0029)	<i>yr1992</i>	-0.0275	(.0008)
	<i>age</i> ²	-0.1880	(.0009)	<i>yr1993</i>	-0.0477	(.0013)
	<i>age</i> ³	0.0277	(.0001)	<i>yr1994</i>	-0.0437	(.0018)
	<i>age</i> ⁴	-0.0016	(.0000)	<i>yr1995</i>	-0.0352	(.0020)
<i>female</i> ×	<i>age</i>	0.0036	(.0007)	<i>yr1996</i>	-0.0225	(.0026)
	<i>age</i> ²	-0.0117	(.0004)	<i>yr1997</i>	0.0036	(.0029)
	<i>age</i> ³	0.0030	(.0001)	<i>yr1998</i>	0.0442	(.0033)
	<i>age</i> ⁴	-0.0002	(.0000)	<i>yr1999</i>	0.0550	(.0037)
<i>black</i> ×	<i>age</i>	-0.0004	(.0012)	<i>yr2000</i>	0.0670	(.0040)
	<i>age</i> ²	-0.0025	(.0007)	<i>yr2001</i>	0.0619	(.0043)
	<i>age</i> ³	0.0005	(.0001)	<i>yr2002</i>	0.0696	(.0046)
	<i>age</i> ⁴	0.0000	(.0000)	<i>yr2003</i>	0.0659	(.0049)
<i>hispanic</i> ×	<i>age</i>	0.0263	(.0008)	<i>yr2004</i>	0.0751	(.0053)
	<i>age</i> ²	-0.0173	(.0009)	<i>yr2005</i>	0.0776	(.0058)
	<i>age</i> ³	0.0029	(.0002)	<i>yr2006</i>	0.0830	(.0062)
	<i>age</i> ⁴	-0.0001	(.0000)	<i>yr2007</i>	0.0927	(.0065)
	<i>sixq2</i>	0.6879	(.0079)	<i>yr2008</i>	0.0880	(.0069)
	<i>sixq3</i>	1.5227	(.0115)	<i>yr2009</i>	0.0784	(.0074)
	<i>sixq4</i>	2.0854	(.0119)	<i>yr2010</i>	0.0843	(.0077)
	<i>sixq5</i>	2.5327	(.0112)	σ	0.3659	(.0003)
	<i>sixq6</i>	2.6913	(.0084)			
	<i>sixqlft</i>	-0.0951	(.0031)			
	<i>sixqrigh</i>	-0.0568	(.0027)			
	<i>sixq4th</i>	0.0989	(.0003)			
	<i>sixqinter</i>	-0.4108	(.0027)			

Table entries are means and Monte Carlo Standard Errors of the parameters on the indicated control variable across 7,922 draws from the Gibbs sampler described in the text.

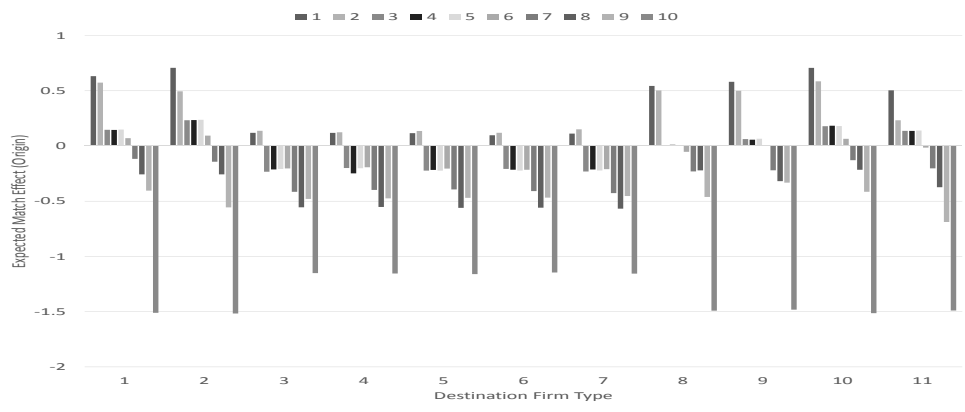
Table OA2: Regression of Structural Wage Decomposition Components on AKM Estimates of Wage Decomposition Components

	$X\beta_{Gibbs}$	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}
$X\beta_{AKM}$	0.4084 (.0021)	0.0357 (.0020)	0.0035 (.0012)	0.4664 (.0125)	0.0047 (.0001)
θ_{AKM}	0.3302 (.0028)	0.1946 (.0099)	0.0265 (.0014)	0.1652 (.0095)	0.0069 (.0002)
ψ_{AKM}	0.2424 (.0011)	0.2398 (.0135)	0.9221 (.0243)	-0.3824 (.0281)	0.0608 (.0009)
μ_{AKM}	0.0384 (.0000)	-0.0020 (.0002)	0.2524 (.0106)	0.5658 (.0111)	0.1432 (.0009)
ε_{AKM}	-0.0243 (.0001)	0.0069 (.0006)	0.0036 (.0008)	0.0351 (.0017)	0.9574 (.0001)
Constant	5.9418 (.0302)	-0.1450 (.0086)	0.0103 (.0006)	-3.4534 (.0838)	-0.0322 (.0006)

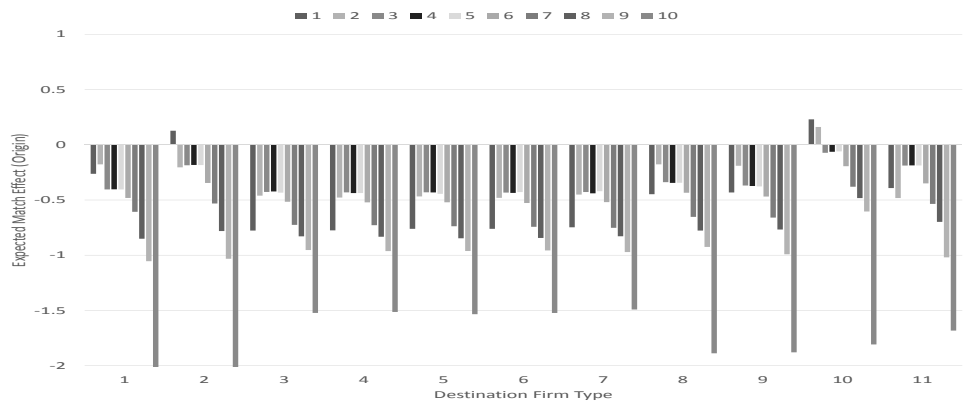
Results from running a regression of the wage components estimated under the endogenous mobility model on wage components estimated using the AKM decomposition. The reported values are the mean parameter estimate and the correlated-draw Monte Carlo standard errors across 7,922 draws from the Gibbs sampler.



(a) Worker Type 1



(b) Worker Type 5



(c) Worker Type 10

Figure OA1: Expected match effect within origin/destination employer type cells, disaggregated by worker type. Legend is firm types.