Online Appendix to Accompany "Modeling Endogenous Mobility in Earnings Determination"

John M. Abowd
U.S. Census Bureau and Department of Economics, Cornell University
and
Kevin L. McKinney
Center for Economic Studies, U.S. Census Bureau
and
Ian M. Schmutte
Department of Economics, University of Georgia

Please consult the published manuscript or pre-print, which is available via http://digitalcommons.ilr.cornell.edu/ldi/28/.

A Supplemental Tables and Figures

Table OA1: Covariance Matrix of Earnings Equation Parameters: LEHD Data 0.5% Sample, $(10, 10, 10)$ Model

	(1) $\ln w$	(2) $X\beta_{AKM}$	(3) θ_{AKM}	(4) ψ_{AKM}	(5) μ_{AKM}	(6) ε_{AKM}	(7) $X\beta_{Gibbs}$	(8) θ_{Gibbs}	(9) ψ_{Gibbs}	(10) μ_{Gibbs}	(11) ε_{Gibbs}
$\ln w$	1.949	0.920	0.629	0.278	0.078	0.056	0.919	0.551	0.327	0.045	0.107
$X\beta_{AKM}$	0.920	1.989	−1.022	0.029	−0.001	0.001	0.693	0.151	0.075	−0.005	0.006
θ_{AKM}	0.629	−1.022	1.599	0.056	0.000	0.000	0.135	0.315	0.112	0.038	0.029
ψ_{AKM}	0.278	0.029	0.056	0.194	0.000	0.000	0.085	0.087	0.126	−0.029	0.010
μ_{AKM}	0.078	−0.001	0.000	0.000	0.080	0.000	0.007	0.000	0.016	0.044	0.011
ε_{AKM}	0.056	0.001	0.000	0.000	0.000	0.056	0.004	0.000	0.000	0.000	0.052
$X\beta_{Gibbs}$	0.919	0.693	0.135	0.085	0.007	0.004	0.701	0.146	0.090	−0.017	0.000
θ_{Gibbs}	0.551	0.151	0.315	0.087	0.000	0.000	0.146	0.707	0.003	−0.305	0.000
ψ_{Gibbs}	0.327	0.075	0.112	0.126	0.016	0.000	0.090	0.003	0.485	−0.251	0.000
μ_{Gibbs}	0.045	−0.005	0.038	−0.029	0.044	0.000	−0.017	−0.305	−0.251	0.619	0.000
ε_{Gibbs}	0.107	0.006	0.029	0.010	0.011	0.052	0.000	0.000	0.000	0.000	0.107

Table entries are means of the covariance between the indicated variables across $9,968$ draws from the Gibbs sampler described in the text using the estimation sample, with number of person-year observations 395,930.

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

	$X\beta_{Gibbs}$	θ_{Gibbs}	ψ_{Gibbs}	μ_{Gibbs}	ε_{Gibbs}
$X\beta_{AKM}$	0.2600	0.0402	0.0129	0.2791	0.0159
	(.0010)	(.0031)	(.0017)	(.0113)	(.0004)
θ_{AKM}	0.2389	0.2094	0.0559	0.2116	0.0265
	(.0021)	(.0105)	(.0022)	(.0071)	(.0008)
ψ_{AKM}	0.3305	0.3812	0.6308	-0.2547	0.0413
	(.0012)	(.0475)	(.0438)	(.0181)	(.0020)
μ_{AKM}	0.0944	0.0006	0.2012	0.5519	0.1374
	(.0005)	(.0001)	(.0114)	(.0137)	(.0028)
ε_{AKM}	0.0637	-0.0007	-0.0002	-0.0048	0.9359
	(.0003)	(.0001)	(.0000)	(.0002)	(.0003)
Constant	7.2656	-0.2421	-0.0344	-2.2883	-0.1258
	(.0574)	(.0186)	(.0015)	(.0774)	(.0035)

Results from running a regression of the earnings components estimated under the endogenous mobility model on earnings components estimated using the AKM decomposition. The reported values are the mean parameter estimate and the correlated-draw Monte Carlo standard errors across 9,968 draws from the Gibbs sampler using the estimation sample, with number of person-year observations 395,930.

Table OA3: Observed Covariate Parameters: AKM and Structural Estimates

	Variable	(1) AKM Model Estimate	(2) Structural Model Mean	(MCSE)
	age	0.5555	0.6321	(.0064)
	age^2	-0.1864	-0.2087	(.0023)
	age^3	0.0277	0.0311	(.0003)
	age^4	-0.0016	-0.0018	(.0000)
$female\times$	age	-0.2534	0.0031	(.0021)
	age^2	0.0866	-0.0091	(.0010)
	age^3	-0.0136	0.0019	(.0002)
	age^4	0.0008	-0.0001	(.0000)
$black\times$	age	-0.2226	0.0157	(.0025)
	age^2	0.0788	-0.0160	(.0019)
	age^3	-0.0127	0.0035	(.0004)
	age^4	0.0008	-0.0002	(.0000)
$hispanic\times$	age	0.0644	0.0369	(.0063)
	age^2	-0.0323	-0.0173	(.0050)
	age^3	0.0065	0.0042	(.0012)
	age^4	-0.0004	-0.0002	(.0001)
	$sixq2$	0.8851	0.7553	(.0069)
	$sixq3$	1.6680	1.6226	(.0088)
	$sixq4$	2.0301	2.1874	(.0115)
	$sixq5$	2.3843	2.6303	(.0120)
	$sixq6$	2.6301	2.7496	(.0103)
	$sixqleft$	-0.1161	-0.0629	(.0037)
	$sixqright$	-0.1054	-0.0301	(.0042)
	$sixq4th$	-0.3500	0.0882	(.0009)
	$sixqinter$	0.2688	-0.3877	(.0030)
	$yr2000$	0.0233	0.0044	(.0007)
	$yr2001$	0.0308	-0.0020	(.0008)
	$yr2002$	0.0494	-0.0089	(.0012)
	$yr2003$	0.0611	-0.0097	(.0014)
	σ		0.3277	(.0039)
Num. Obs.		60,123,894	395,930	

Table entries are parameters on time-varying characteristics included in both the AKM model and the structural endogenous mobility model. Column (1) reports parameter estimates from the fit of the AKM model to the LEHD analysis population. Columns under (2) report the posterior means and Monte Carlo Standard Errors for parameters on the indicated control variable based on 9,968 draws from the Gibbs sampler.

Table OA4: Serial correlation in structural residuals

	Correlation Coeff.	MCSE
$\rho_{t,t-1}$	0.1153	0.0054
$\rho_{t,t-2}$	-0.0294	0.0060
$\rho_{t,t-3}$	-0.0805	0.0067

Each row reports the posterior mean and MCSE across 9,968 draws from the Gibbs sampler of the within-worker correlation in the unexplained residual portion of earnings from the endogenous mobility model.

Table OA5: Mean Log Earnings Net of Observed Characteristics for Workers Who Change Jobs in 2001 by Quartile of Firm Effect for Origin and Destination Firms, 1999–2003

Transition Cell	1999	2000	2001	2002	2003
1 to 1	0.587	0.604	0.642	0.645	0.647
1 to 2	0.625	0.614	0.975	1.017	1.041
1 to 3	0.726	0.703	1.353	1.387	1.411
1 to 4	0.799	0.746	1.803	1.821	1.853
2 to 1	1.017	1.030	0.756	0.765	0.759
2 to 2	1.208	1.239	1.264	1.264	1.273
2 to 3	1.354	1.356	1.585	1.602	1.622
2 to 4	1.496	1.488	1.996	2.013	2.041
3 to 1	1.449	1.437	0.741	0.770	0.760
3 to 2	1.567	1.577	1.367	1.394	1.398
3 to 3	1.806	1.834	1.857	1.845	1.860
3 to 4	1.997	2.004	2.205	2.205	2.236
4 to 1	2.025	2.042	0.727	0.723	0.717
4 to 2	2.025	2.045	1.378	1.413	1.430
4 to 3	2.124	2.182	1.964	1.960	1.967
4 to 4	2.403	2.491	2.487	2.443	2.456

The table entries are means of log earnings net of the effect of observed time-varying characteristics for a specific year and transition cell. The sample is the LEHD analysis population described in Section 5.1.1 who change jobs exactly once between 1999 and 2003, and where the year of job transition is 2001. This sample follows 566,300 workers across 183,100 unique firms for a total of 2,832,000 person-year observations. Each job is assigned to a quartile based on the estimated AKM firm effect. The "Transition Cell" column indicates the quartile of the origin and destination job. Figure 4 displays a selection of these transition summaries.

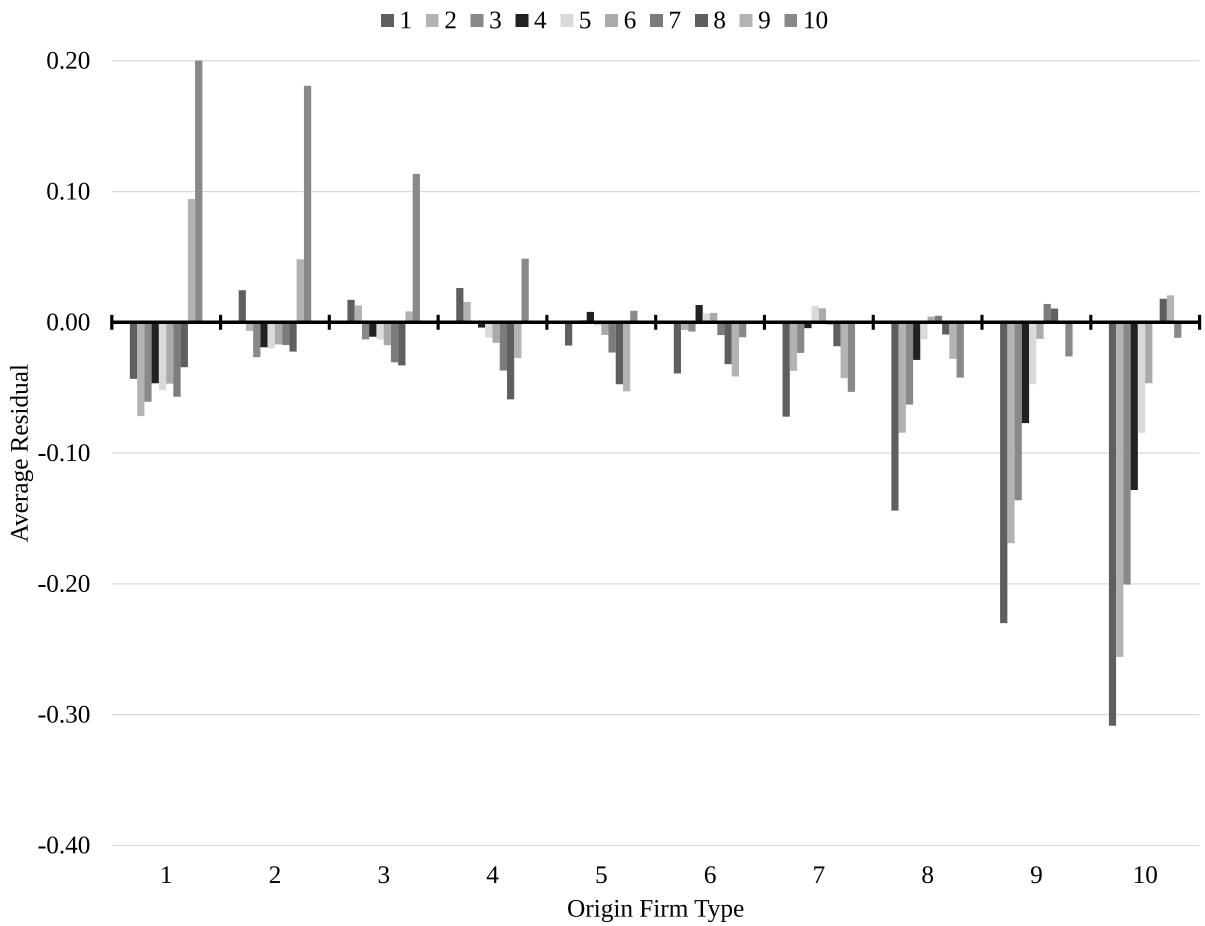
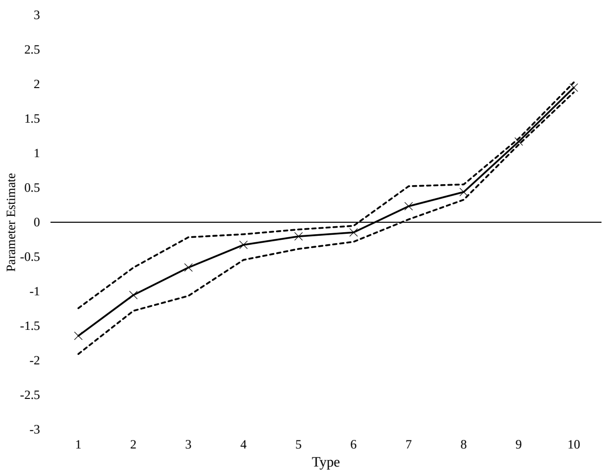
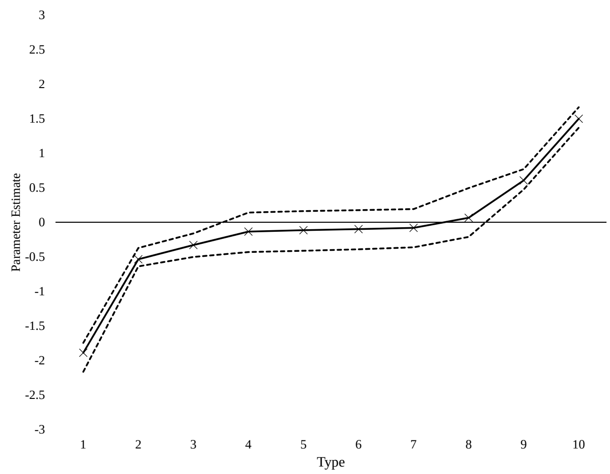


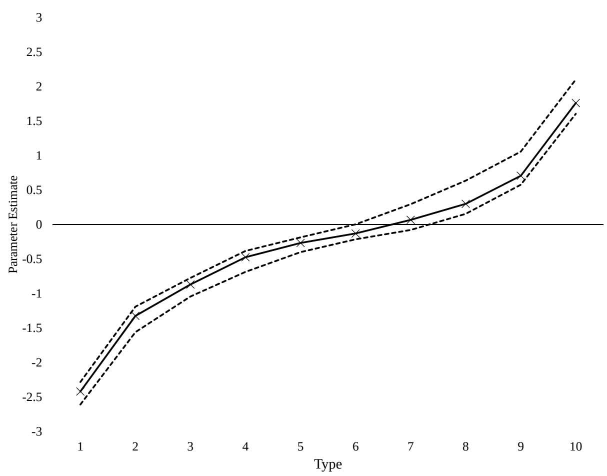
Figure OA1: Mean change in the AKM residual within origin/destination firm effect decile. Legend is for destination firm types.



(a) Structural Worker Effect, θ



(b) Structural Firm Effect, ψ



(c) Structural Match Effect, μ

Figure OA2: Posterior distribution of wage equation parameters. The solid line indicates the posterior mean. Dashed lines indicate the 5th and 95th percentiles.

B Formal Test of Endogenous Mobility

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.

B.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

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

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

B.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 $\overline{\varepsilon}_{it-1}$ for those individuals who change employers between periods $t-1$ and t. Formally,

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

An individual for whom $\overline{\varepsilon}_{it-1} > 0$ 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 $\overline{\varepsilon}_{it-1} < 0$.

B.2.1 Derivation of the Match Effects Test Statistic

To form a test statistic that captures the potential for $\overline{\varepsilon}_{it-1}$ 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 $\overline{\varepsilon}_{it-1}$:

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

The joint probability of observing n_{abcd} is

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

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

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

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

$$X^2_{\nu_1} = Test\left(\pi_{abcd} = \pi_{abc+}\pi_{+++d}\right) \tag{B-2}$$

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

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

B.2.2 Computation of the Match Effects Test

We compute the test statistic (B-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 $t = 1999, ..., 2003$. 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 $\overline{\varepsilon}_{it-1}$ distributions. Then the counts (B-1) were tabulated using all observations in the job-changer population and used to compute the relevant marginal frequencies for the test.

B.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

$$\widetilde{\varepsilon}_{jt} = \frac{\sum\limits_{\{i|J(i,t)=j\}} \widehat{\varepsilon}_{it}}{\sum 1\left\{i|J\left(i,t\right)=j\right\}}.$$

An employer for whom $\widetilde{\varepsilon}_{jt} > 0$ has paid higher than expected wages in period t; and the opposite is true for $\widetilde{\varepsilon}_{jt} < 0$. Although there could be many reasons for this, we will refer to $\widetilde{\varepsilon}_{jt}$ as a measure of workforce productivity. However, the exogenous mobility hypothesis is silent about the meaning of $\widetilde{\varepsilon}_{jt}$. What matters is its relationship to the within-employer distribution of θ_i. If $\widetilde{\varepsilon}_{js}$ 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

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

and

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

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

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

and

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

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

$$X^2_{\nu_2} = Test\left(\ln\left(\frac{\pi_{abc|s}}{\pi_{abc|t}}\right) = \ln\left(\frac{\pi_{ab+|s}}{\pi_{ab+|t}}\right)\right) \tag{B-3}$$

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

B.3.1 Derivation of the Productive Workforce Test Statistic

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

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

where the notation is as follows:

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

The change in main effects of $Q\left(\psi_j\right)$ from period s to t, $\left(\mu_{a|s}-\mu_{a|t}\right)$, must be 0 since the population of employers is restricted to be identical in both periods. Similarly, the change in main effects of $Q\left(\widetilde{\varepsilon}_{js}\right)$, $\left(\mu_{c|s}-\mu_{c|t}\right)$, must be 0 since the workforce productivity distribution is only measured at period s. The change in interaction of $Q\left(\psi_j\right)$ and $Q\left(\widetilde{\varepsilon}_{js}\right)$, $\left(\gamma_{ac|s}-\gamma_{ac|t}\right)$, 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 $Q\left(\theta_i\right)$, $\left(\mu_{b|s}-\mu_{b|t}\right)$, with $df=\left(\#Q\left(\theta_i\right)-1\right)$ and the change in interaction of $Q\left(\psi_j\right)$ and $Q\left(\theta_i\right)$, $\left(\gamma_{ab|s}-\gamma_{ab|t}\right)$, with $df=\left(\#Q\left(\psi_j\right)-1\right)\times\left(\#Q\left(\theta_i\right)-1\right)$. The parameters affected by the null hypothesis are the change in interaction of $Q\left(\theta_i\right)$ and $Q\left(\widetilde{\varepsilon}_{js}\right)$, $\left(\gamma_{bc|s}-\gamma_{bc|t}\right)$, with $df=\left(\#Q\left(\theta_i\right)-1\right)\times\left(\#Q\left(\widetilde{\varepsilon}_{js}\right)-1\right)$ and the change in interaction of $Q\left(\psi_j\right),Q\left(\theta_i\right)$ and $Q\left(\widetilde{\varepsilon}_{js}\right)$, $\left(\rho_{abc|s}-\rho_{abc|t}\right)$, with $df=\left(\#Q\left(\psi_j\right)-1\right)\times\left(\#Q\left(\theta_i\right)-1\right)\times\left(\#Q\left(\widetilde{\varepsilon}_{js}\right)-1\right)$. Under the null hypothesis $\left(\gamma_{bc|s}-\gamma_{bc|t}\right)=0$ and $\left(\rho_{abc|s}-\rho_{abc|t}\right)=0$ with $df=\nu_2=\left(\#Q\left(\theta_i\right)-1\right)\times\left(\#Q\left(\widetilde{\varepsilon}_{js}\right)-1\right)+\left(\#Q\left(\psi_j\right)-1\right)\times\left(\#Q\left(\theta_i\right)-1\right)\times\left(\#Q\left(\widetilde{\varepsilon}_{js}\right)-1\right)$.

B.3.2 Computation of the Productive Workforce Test Statistics

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

$$x_j=\left[\begin{array}{c}\frac{n_{j1t}}{n_{j+t}}-\frac{n_{j1s}}{n_{j+s}}\\ \frac{n_{j2t}}{n_{j+t}}-\frac{n_{j2s}}{n_{j+s}}\\ \cdots\\ \frac{n_{j(\#Q(\theta_i)-1)t}}{n_{j+t}}-\frac{n_{j(\#Q(\theta_i)-1)s}}{n_{j+s}}\end{array}\right]$$

where

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

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

$$\bar{x}_{ac} = \frac{\sum\limits_{\left\{j|Q(\psi_j)=a\wedge Q(\tilde{\varepsilon}_{js})=c\right\}} n_{j+s}x_j}{\sum\limits_{\left\{j|Q(\psi_j)=a\wedge Q(\tilde{\varepsilon}_{js})=c\right\}} n_{j+s}}$$

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

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

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

$$\bar{x}_{a} = \frac{\sum\limits_{\left\{j|Q(\psi_j)=a\right\}} n_{j+s}x_j}{\sum\limits_{\left\{j|Q(\psi_j)=a\right\}} n_{j+s}}.$$

Then,

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

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

C Posterior Distribution of the Parameter Vector

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

$$\begin{aligned} p\left(\rho|Y,Z\right) \quad &\propto \quad \pounds\left(\rho|Y,Z\right)\frac{1}{\sigma^{\nu_0+1}}\exp\left(-\frac{s_0^2}{\sigma^2}\right)\prod_{\ell=1}^{L}\pi_{a\ell}^{\frac{1}{L}-1}\prod_{m=1}^{M}\pi_{bm}^{\frac{1}{M}-1} \qquad \text{(C-1)} \\ &\times\prod_{\ell=1}^{L}\prod_{m=0}^{M}\prod_{q=1}^{Q}\left(\pi_{q|\ell m}^{\frac{1}{Q}-1}\gamma_{\ell mq}^{\frac{1}{2}-1}\left(1-\gamma_{\ell mq}\right)^{\frac{1}{2}-1}\prod_{m'=0}^{M}\delta_{m'|\ell mq}^{\frac{1}{M+1}-1}\right). \end{aligned}$$

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

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

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

The posterior distribution of the wage equation parameters is

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

where

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

and

$$\sigma^2 \sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{2}{\nu s^2}\right). \tag{C-4}$$

The posterior distributions for the latent heterogeneity types are Dirichlet:

$$\pi_a \sim \mathrm{D}\left(n_{a1} + \left(\frac{1}{L}\right), \ldots, n_{aL} + \left(\frac{1}{L}\right)\right); \tag{C-5}$$

$$\pi_b \sim \mathrm{D}\left(n_{b1} + \left(\frac{1}{M}\right), \ldots, n_{bM} + \left(\frac{1}{M}\right)\right); \tag{C-6}$$

$$\pi_{k|ab} \sim \mathrm{D}\left(n_{k|ab1} + \left(\frac{1}{Q}\right), \ldots, n_{k|abQ} + \left(\frac{1}{Q}\right)\right). \tag{C-7}$$

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

$$\gamma_{lmq} \sim \mathrm{D}\left(n_{\ell mq}^{sep} + \left(\frac{1}{2}\right), n_{\ell mq}^{stay} + \left(\frac{1}{2}\right)\right); \tag{C-8}$$

$$\delta_{b|lmq} \sim \mathrm{D}\left(n_{0|\ell mq}^{trans} + \left(\frac{1}{M+1}\right), \ldots, n_{M|\ell mq}^{trans} + \left(\frac{1}{M+1}\right)\right). \tag{C-9}$$

D 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

$$\Lambda = \mathrm{diag}([\lambda_{111}, \lambda_{112}, \ldots, \lambda_{LMQ}]^T). \tag{D-1}$$

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

$$\tilde{y}_{\ell,m,q} = \alpha + \theta_\ell + \psi_m + \mu_q. \tag{D-2}$$

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

$$\begin{bmatrix} \tilde{D} & \tilde{F} & \tilde{G} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}. \tag{D-3}$$

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

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

and the full cross-product matrix is

$$\begin{bmatrix} \tilde{D} & \tilde{F} & \tilde{G} \end{bmatrix}^T \Lambda \begin{bmatrix} \tilde{D} & \tilde{F} & \tilde{G} \end{bmatrix} = \begin{bmatrix} \tilde{D}^T \Lambda \tilde{D} & \tilde{D}^T \Lambda \tilde{F} & \tilde{D}^T \Lambda \tilde{G} \\ \tilde{F}^T \Lambda \tilde{D} & \tilde{F}^T \Lambda \tilde{F} & \tilde{F}^T \Lambda \tilde{G} \\ \tilde{G}^T \Lambda \tilde{D} & \tilde{G}^T \Lambda \tilde{F} & \tilde{G}^T \Lambda \tilde{G} \end{bmatrix}. \tag{D-5}$$

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

E Estimation and Data Details

E.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 (13). 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, $\mathrm{N}(j)$, is the set of all employers, j', that are degree-one connected to j. Equation (13) implies that if j'' is not in $\mathrm{N}(j)$ and j is not in $\mathrm{N}(j'')$, 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 types 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.

E.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 E.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.3 Details of Variable Construction

Here we describe how the analysis variables are constructed from the LEHD microdata. Note that the raw UI records that supply the LEHD infrastructure are quarterly earnings records. Our analysis data are at the job-year level. Our key dependent variable, annual earnings, is constructed by summing all quarterly earnings records (converted to base year 2000 dollars using CPI-U) for the same job (worker matched to firm) over the year in question. To deal with outliers in earnings, we Winsorize at the 0.01 and 99.99 percentiles.

Demographic characteristics of the worker are linked from the National Individual Characteristics File (NICF). These characteristics originate from Social Security records and other sources, including Census 2000 and the American Community Survey. These characteristics, and their construction, are described in Abowd et al. (2009), and subsequent internal research.

We also construct controls for the sequence of earnings records observed over the year. These controls address the problem that jobs that end mid-year will mechanically have lower earnings than jobs that last all year, even if the rate at which labor market earnings are acquired remains constant. These are based on a bit string, called `sixqwindow`, that records the observed pattern of quarters with positive earnings for a given job-year combination. The variable `sixqwindow` also records whether the worker was observed employed on the current job at the end of the preceding year and in the start of the subsequent year. So, for example, `sixqwindow`=000011 for a job that started in the last quarter of the current year and continued into the first quarter of the next year. Likewise, `sixqwindow`=110000 for a job on which a worker was employed at the end of the last year, and also reported earnings into the first quarter of the current year. As a final example, a job which is continuing from the preceding year, and continuing into the next year, and in which the worker was employed all year will have `sixqwindow`=111111.

There are 60 feasible values of `sixqwindow` (since strings of the form x0000y are ruled out). We summarize the salient information from these strings in a set of ten indicator variables. These are

- `sixq1` is an indicator equal to 1 if the sum of entries in `sixqwindow` is equal to 1, and zero otherwise.
- `sixq2` is an indicator equal to 1 if the sum of entries in `sixqwindow` is equal to 2, and zero otherwise.
- `sixq3--sixq6` are defined equivalently to the above.
- `sixqleft` is equal to 1 if `sixqwindow` has a continuous list of zeros from the right and ones from the left.
- `sixqright` is equal to 1 if `sixqwindow` has a continuous list of ones from the right and zeros from the left.
- `sixqinter` is equal to 1 if `sixqwindow` has a continuous list of ones interrupted by a single sequence of zeros.
- `sixq4th` is equal to 1 if `sixqwindow` indicates the worker was employed in the fourth quarter of the year.

These variables are effective in dealing with differences in job attachment over the year. Another option is to use the same information to convert the earnings data into an annualized level that measures the earnings a worker would earn if they accrued earnings at the same rate through the entire year. The latter approach requires ad hoc assumptions about when jobs tend to end within the quarter. By contrast, we can see exactly how the endogenous mobility model treats such cases from the coefficient estimates in Table OA3. For example, consider the contrast between a job-year observation for which the worker is continuously employed relative to a job-year observations for which the worker is employed through the third quarter. In terms of the estimated parameters, the worker employed full time full year will have sixq6=1, sixqleft=1, sixqright=1 (and all other job attachment dummies=0). By contrast, the worker employed through the second quarter will have sixq3=1 and sixqleft=1 (and all other attachment dummies=0). Assuming all else is the same, the difference in predicted log earnings is 0.5321. That is, earnings on a job that ends during the third quarter are predicted to be 60 percent as large a job that lasts all year and continues into the next. The 60 percent figure captures the fact that jobs with reported earnings in the third quarter likely do not last to the end of the quarter.

E.4 Details of Selecting the Number of Latent Types

Following the guidance in Gelman, Carlin, Stern, Dunson, Vehtari and Rubin (2013, p. 536), our intention was to select a specification in which the number of latent types is an upper bound and let the data reveal the number of occupied types. To that end, through an initial model selection step, we selected the number of latent worker, firm, and match types to make the unexplained variation in the structural earnings model as close as possible to the residual variance from the AKM decomposition. We also favored specifications that are a priori symmetric in the number of latent types. This process resulted in our preferred specification with ten latent types for each dimension of heterogeneity. As discussed in the text, it appears ex post that while there is mass in each of the latent worker and match types, only four of the latent firm types have support. The missing categories are effectively collapsed in the posterior summaries.

An alternative procedure is to use a formal model selection criterion. For non-singular models like ours, the literature cautions that the Akaike, Bayesian, and Deviance Information Criteria are either technically infeasible, or not theoretically well-justified (Watanabe 2013; Gelman, Carlin, Stern, Dunson, Vehtari and Rubin 2013). The recommended alternative is the Watanabe-Akaike Information Criterion (WAIC) as it is both fully Bayesian, computationally tractable, and formally connected to cross-validation. See Gelman, Hwang and Vehtari (2013) for complete details.

We applied the WAIC to models with 3, 5, 7, and 10 latent types on each dimension. For the model selection exercise, we draw 1000 samples under each specification and compute the WAIC based on a 1-in-25 thinned subsample after a 500 sample burn-in. The results appear in Table OA6.

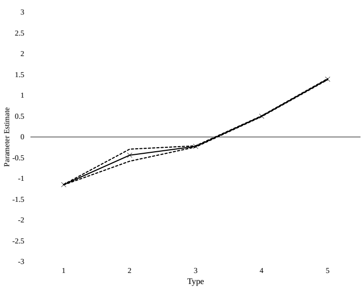
The WAIC is lowest for the model with 5 latent heterogeneity types and highest for our preferred model with 10 latent types. The model with 10 types has non-trivially higher likelihood than the others. These findings are consistent with our view that the model with 10 latent types represents an upper bound on the number for our purpose of fitting the existing data.

Figures OA3 and OA4 show the estimated wage components and distribution across latent types respectively for the models with 5, 7, and 10 latent types. These figures indicate that the pattern of model estimates is broadly consistent across specifications. It is only when we go to the model with 10 types that it becomes evident that several of the latent firm types are not filled. Again, these categories are effectively collapsed when the data are postprocessed to generate posterior

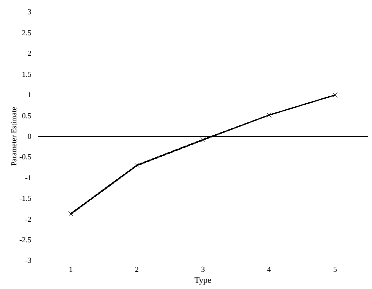
Table OA6: Likelihood and Watanabe-Akaike Information Criterion under Different Specifications

Model	WAIC	Likelihood
$L, M, Q = 3$	$1,174,843$	$-431,720$
$L, M, Q = 5$	$1,109,066$	$-363,529$
$L, M, Q = 7$	$1,139,649$	$-367,601$
$L, M, Q = 10$	$1,153,349$	$-337,972$

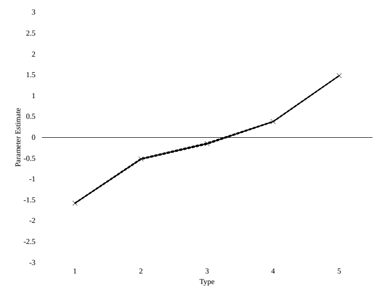
summaries. To instead perform an exhaustive model selection search using WAIC would require us to separately fit the model under each of the 1000 possible combinations of types. It is therefore infeasible. Given the goals of our analysis, the methodological literature and the data both support our choice to model 10 latent types as an explicit upper bound.



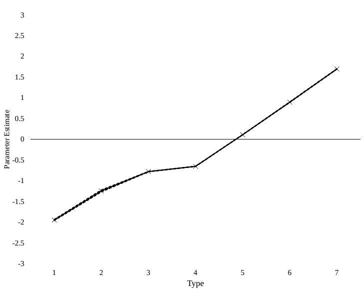
(a) Structural Worker Effect, θ



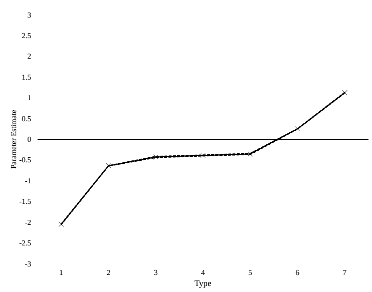
(b) Structural Firm Effect, ψ



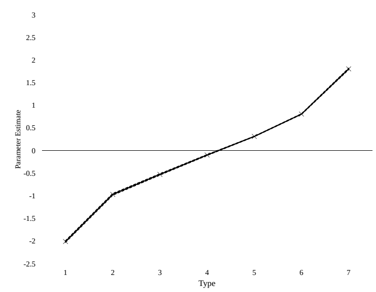
(c) Structural Match Effect, μ



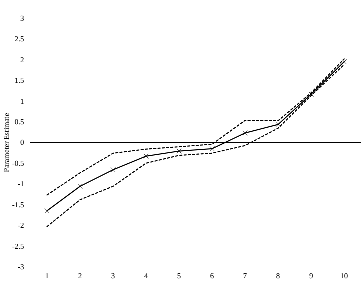
(d) Structural Worker Effect, θ



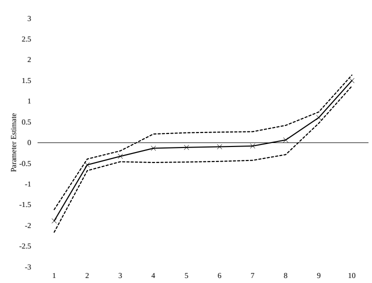
(e) Structural Firm Effect, ψ



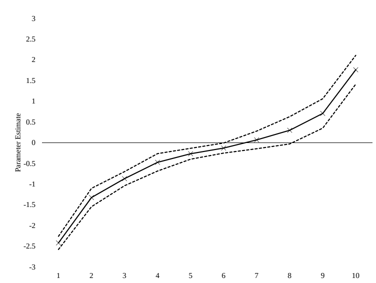
(f) Structural Match Effect, μ



(g) Structural Worker Effect, θ

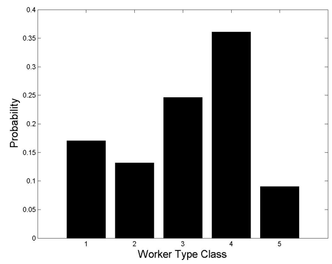


(h) Structural Firm Effect, ψ

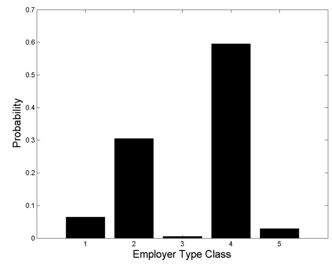


(i) Structural Match Effect, μ

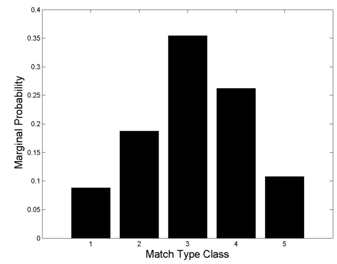
Figure OA3: Posterior distribution of earnings equation parameters. Dashed lines indicate $\pm 2 \times$ MCSE. The top row reports the specification with $L = M = Q = 5$ latent types. The second row reports the specification with $L = M = Q = 7$ latent types. The second row reports the specification with $L = M = Q = 10$ latent types.



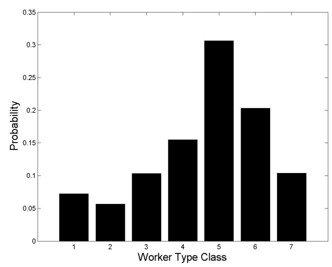
(a) Dist. of Worker Types



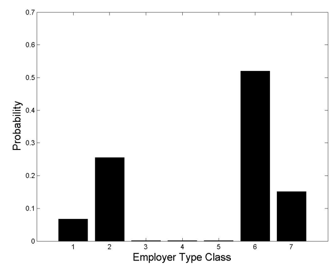
(b) Dist. of Employer Types



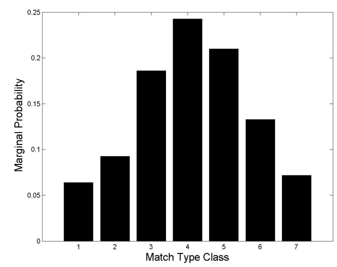
(c) Marginal Dist. of Match Types



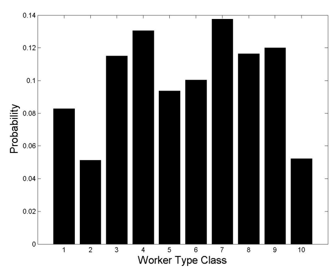
(d) Dist. of Worker Types



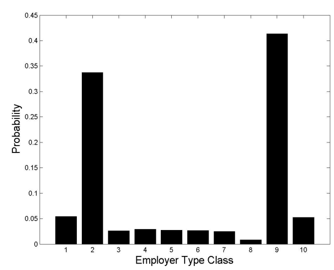
(e) Dist. of Employer Types



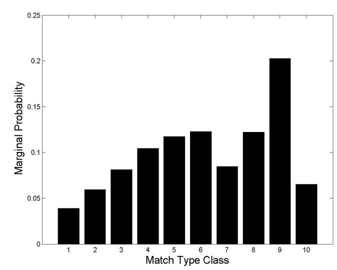
(f) Marginal Dist. of Match Types



(g) Dist. of Worker Types



(h) Dist. of Employer Types

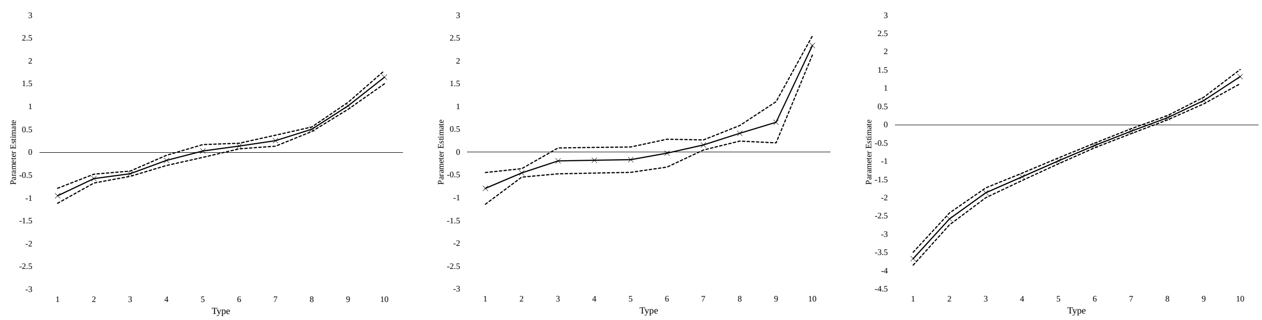


(i) Marginal Dist. of Match Types

Figure OA4: Posterior distribution of workers, employers, and matches across latent types. The top row reports the specification with $L = M = Q = 5$ latent types. The second row reports the specification with $L = M = Q = 7$ latent types. The second row reports the specification with $L = M = Q = 10$ latent types.

F Results using Extended Work Histories

We report results of estimating the structural model on an extended version of our main analysis sample. Specifically, we augment the main analysis sample by attaching the complete work history from 1990–2010 for each of the workers. This "extended work history" sample includes 1,778,490 person-year observations that cover 181,592 firms, and 389,718 matches. It is constructed to contain the primary analysis sample from the main text as a strict subset. The results reported here are based on $7,922$ draws from three parallel runs of the Gibbs sampler after removing a 300 iteration burn-in. A complete archive of these results is available by request.



(a) Structural Worker Effect, θ (b) Structural Firm Effect, ψ (c) Structural Match Effect, μ

Figure OA5: Posterior distribution of earnings equation parameters. Dashed lines indicate the region within $2 \times MCSE$ of the estimate.

Table OA7 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.

Table OA8 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). It is the analogue to Table 2 from the main text.

Table OA9 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. It is the analogue to Table 3 from the main text.

Table OA7: Posterior Distribution of Worker, Firm, and Match Population Heterogeneity: Extended Work Histories

	(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

Results from the strucutral model estimated using extended work histories. It is structured identically to Table 1 from the main text.

Table OA8: Correlation Matrix of Earnings Equation Parameters: Extended Work Histories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	$\ln w$	$X\beta_{AKM}$	θ_{AKM}	ψ_{AKM}	μ_{AKM}	ε_{AKM}	$X\beta_{Gibbs}$	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 OA9: Regression of Log Revenue Per Worker on Structural and AKM Estimates of Earnings Decomposition Components: Extended Work Histories

	(1) Structural		(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. $X\beta$	0.0343	0.0063	-0.0231	0.0045
Intercept	3.4898	0.0532	3.9476	0.0247
N	$60,116$		$60,116$	

Bibliography

Abowd, J. M., Stephens, B. E., Vilhuber, L., Andersson, F., McKinney, K. L., Roemer, M. and Woodcock, S. (2009)

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.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Gelman, A., Hwang, J. and Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models, *Statistics and Computing* **24**(6): 997–1016.

Geyer, C. (2011). Introduction to Markov Chain Monte Carlo, *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, Chapman and Hall/CRC.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo, *Statistical Science* **7**(4): pp. 473–483.

Kosorok, M. R. (2000). Monte carlo error estimation for multivariate markov chains, *Statistics & Probability Letters* **46**(1): 85 – 93.

Watanabe, S. (2013). A widely applicable Bayesian information criterion, *Journal of Machine Learning Research* **14**(1): 867–897.