The Anatomy of Sorting: Evidence from Danish Data

Rasmus Lentz[1]   Suphanit Piyapromdee[2]   Jean-Marc Robin[3]

[1]UW-Madison & The Dale T. Mortensen Centre, [2]UCL & IFS, [3]Sciences Po & UCL

BOT 2018

INTRODUCTION

Introduction

1. What is the assignment of workers to firms with respect to **unobserved characteristics**?

2. How is it realized?

- The second question is new.

- The current literature measures sorting, but says very little on the matching process itself.

Measuring sorting

- We need a statistical model.
- Descriptive approaches:
  - Fixed effects, Abowd, Kramarz, Margolis (ECMA, 99)
  - Discrete types, Bonhomme, Lamadon, Manresa (2017)
- Structural approaches (Lise et al. 2015, Lise, Robin 2017, Hagedorn et al. 2017, Lopez de Melo 2018, Bagger and Lentz 2018) constrain the matching mechanism.
- In this paper we want to be as general as possible on the way unobserved heterogeneity separately conditions wages and mobility => BLM

AKM

- Additive, fixed effect model of wages:

$$\ln w_{it} = \alpha_i + \psi_j + x_{it}\beta + \varepsilon_{it}$$

- No restriction on matching.
- Model is estimated by OLS on matched employer-employee data. Not many job transitions per worker $\Rightarrow$ small-T biases.
  - Finite-sample bias correction of Andrews, Gil, Schank, and Upward (JRSS, 2008) changes estimates moderately.
- No non-linear version of AKM.

BLM

- Long tradition of discrete heterogeneity in economics since Heckman and Singer (ECMA, 1984).
- BLM assume discrete worker types k and firm types ℓ, and unrestricted $F_{k\ell}, M_{k\ell\ell'}$ (wage distribution and transition probabilites)

1. A classification of firms into discrete classes is first obtained based on within-firm wage distributions (k-means).
2. Conditional on a firm classification, they show that the finite mixture over worker types is identified for fixed T as short as $T = 2$.

Our model

- We use BLM's discrete-type approach,
  - With a special parametric specification of transition probabilities to facilitate interpretation and augment efficiency.
  - With an iterative, more efficient firm classification procedure.
- Transition from prototype to large-scale production: all Danish register data from 1987-2013.
  - Flexible incorporation of observed heterogeneity.
  - EM algorithm design contribution (M step) for estimation of nonlinear mobility parameters.
  - Numerical implementation: Efficient computation delegation in parallel (data structure design and delegation).
- We assume that workers and firms keep the same type for ever but link between wages/mobility and types can change flexibly in observable characteristics (calendar time, experience, tenure).

Main findings 1

- Correlation between worker and firm types is estimated around 28%, stable over time.
- Log-wage variance decomposition:
  - Residual 50%
  - Worker effect 30%
  - Firm effect 5%
  - Covariance 7%
  - Match-specific (ie non linearities in link to wkr and firm types) 5%
  - Observed (tenure and experience) 3%

Main findings 2

- Surprisingly, we find that overall sorting level is largely already obtained at the outset of workers' careers.
- Mobility strengthens sorting during the first 10 years of worker's careers.
- Highly experienced workers move significantly less, locking in place position on ladder. But when they do move, mobility tends to weaken sorting in the latter part of worker's careers.
- Strong sorting on extensive margin.
- The main parameters here are
  - the preference for the job (impacts both direction and frequency of moves),
  - mobility patterns out of non-employment.

THE MODEL

Data structure

- Workers: $i \in \{1, ..., N\}$.
- Firms: $j \in \{1, ..., J\}$. $j = 0$ denotes non-employment.
- t denotes the observation occurrence (week).
- Observations (w_{it}, j_{it}, x_{it}), $t = 1, 2, \ldots, T_i$.
  - $j_{it} \in \{0, 1, ..., J\}$ is the ID of the firm employing worker i in week t.
  - x_{it} are calendar time, potential experience, and tenure.
  - w_{it} is worker i's log-wage rate at time t
- We also observe gender and education (z_i).
- Firms differ in observable characteristics, e.g. private/public sector (ζ_j).

Unobserved heterogeneity

- Firms clustered into L groups indexed by $\ell \in \{0, 1, \ldots, L\}$. Non-employment is $\ell = 0$.
    - Unobserved firm types ℓ treated as fixed effects.
    - Let $F = (\ell_1, \ldots, \ell_J)$ denote a given firm classification.
- Workers clustered into K groups indexed by $k \in \{1, \ldots, K\}$.
    - Unobserved worker types k treated as random effects.
- Workers' and firms' types constrained to remain the same throughout 27 year period.

Likelihood

Lots of conditional independence:

$$L(\beta, F) = \sum_{k=1}^{K} \frac{\pi_k(z_i)\, m_{k,\ell_{i1}}^0(x_{i1})}{q_{\ell_{i1}}(F)} \times \prod_{t=1}^{T} f_{k\ell_{it}}(w_{it}|x_{it})$$
$$\times \prod_{t=1}^{T-1} M_{k\ell_{it}\neg}(x_{it})^{1-D_{it}} \left(\frac{M_{k\ell_{it}\ell_{i,t+1}}(x_{it})}{q_{\ell_{i,t+1}}(F)} \right)^{D_{it}}$$

where

- $\pi_k(z_i)$ is probability of worker type given education and gender
- $m_{k,\ell_{i1}}^0(x_{i1})$ is initial matching probability given current date
- $f_{k\ell_{it}}$ is wage distribution given match type
- $M_{k\ell_{it}\ell_{i,t+1}}$ is transition probability and $M_{k\ell_{it}\neg}$ is probability of staying given current date, tenure and experience
- $q_\ell(F) = \#\{j : \ell_j = \ell\}/J$ is the share of type-ℓ firms (so $1/q_\ell(F)$ is proportional to the probability that this particular firm j_{it} be drawn)

Empirical specification – Wages

- $f_{k\ell}(w|x)$ denotes the wage density, conditional on worker type k and employer type ℓ.
- We use a lognormal distribution,

$$f_{k\ell}(w|x) = \frac{1}{w} \frac{1}{\sigma_{k\ell}(x)} \varphi\left(\frac{\ln w - \mu_{k\ell}(x)}{\sigma_{k\ell}(x)}\right), \quad \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Empirical specification – Transition probabilities

- Probability that type k worker moves from ℓ to ℓ' firm type,

$$M(\ell'|k,\ell,x) \equiv M_{k\ell\ell'}(x) = \lambda_k(x)\nu_{\ell'}(x)P_{k\ell\ell'}(x)$$

(x is date, tenure and experience)
  - λ_k: worker search intensity
  - $\nu_{\ell'}$: sampling rate of firm types (sums to one)
  - $P_{k\ell\ell'}$: Probability that ℓ to ℓ' transition is executed.

- We assume a Bradley-Terry specification, with $P_{k00} = 0$ and

$$P_{k\ell\ell'}(x) = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)}, \quad \sum_{\ell=1}^{L}\gamma_{k\ell} = 1,$$

where $\gamma_{k\ell}$ measures the quality of the match (k,ℓ).

- $M(\neg|k,\ell,x) \equiv M_{k\ell\neg}(x) = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}$ is probability of staying with same employer.

- u-e and e-u transition probabilities completely unrestricted:

$$M(\ell'|k,0,x) = \psi_{k\ell'}(x), \quad M(0|k,\ell,x) = \delta_{k\ell}(x).$$

EM algorithm given firm classification

Iterate the following two steps:

1. For a given firm classification $F^{(s)} = (\ell_1^{(s)}, ..., \ell_J^{(s)})$, find $\beta^{(s)}$ the ML estimate of $\beta = (\underbrace{\mu, \sigma}_{\text{wages}}, \underbrace{\lambda, \nu, \gamma, \delta,}_{\text{mobility}} \underbrace{\psi, \pi, m^0}_{\text{initial}})$ using the EM algorithm.

2. Update firm classification by maximizing the expected log likelihood given observations and current values $F^{(s)}$ and $\beta^{(s)}$

-> Details on CEM algorithm

DATA AND ESTIMATION

Data

- Danish matched employer-employee data, 1987-2013.
  - 5mil workers and 600k firms.
- All matches with average of less than 25 hours per week coded as non-employment.
- Wages reported at annual frequency.
- Worker mobility data reported weekly.
- Time-invariant characteristics z_i :
  - Education: high ($>$ 12 yrs), medium ($=$ 12 yrs), low ($<$ 12 yrs).
  - Gender: male, female
- Time-variant characteristics x_{it} :
  - 9 periods: 87-89, 90-92, ..., 11-13
  - Experience: <5, 5-10, 10-15, >15 years
  - Tenure: short ($<$ 26 weeks), long ($>$ 26 weeks)

Number of groups and labeling

- We set $K = 14$ worker types, $L = 24$ firm types (about the maximum K, L that we can handle with current algorithm and computing facilities).
- Linear projection,

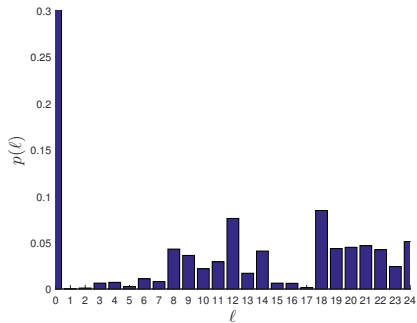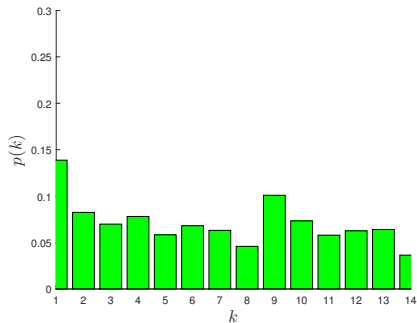$$\mu_{k\ell}(x) = \overline{\mu}(x) + a_k + b_\ell + \widetilde{\mu}_{k\ell}(x),$$

where $\overline{\mu}(x)$ contains tenure*experience interactions.

- Order worker types by a_k.
- Order firm types by b_ℓ.

HETEROGENEITY DISTRIBUTION

Worker k types and firm ℓ types

Distribution of worker types roughly uniform but not firm types

$$p(k) = \frac{1}{N} \sum_i p_i(k) \qquad p(\ell) = \frac{\#\{i, t : \ell_{it} = \ell\}}{NT}$$

Worker *k* types by education and gender

Worker type	Gender		Education		
k	Male	Female	Low	Medium	High
1	0.48	0.52	0.64	0.31	0.05
2	0.13	0.87	0.40	0.52	0.08
3	0.31	0.69	0.28	0.55	0.17
4	0.42	0.58	0.23	0.57	0.20
5	0.45	0.55	0.30	0.45	0.25
6	0.51	0.49	0.18	0.55	0.27
7	0.40	0.60	0.13	0.42	0.45
8	0.62	0.38	0.24	0.44	0.32
9	0.77	0.23	0.27	0.58	0.15
10	0.76	0.24	0.14	0.66	0.20
11	0.55	0.45	0.07	0.33	0.60
12	0.68	0.32	0.07	0.37	0.56
13	0.78	0.22	0.04	0.28	0.68
14	0.83	0.17	0.03	0.20	0.77
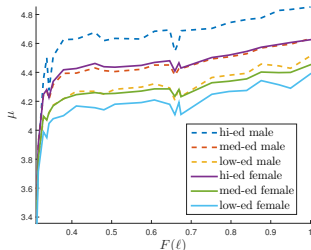
Expected wage $\mu_{k\ell}(x)$ by gender and education

Sorting across k and ℓ

- Copula $c[F(k), F(\ell)] \equiv \frac{p(k,\ell)}{p(k)p(\ell)}$ measures distance of matching proba $p(k, \ell) \equiv \frac{1}{NT} \sum_{i,t:\ell_{it}=\ell} p_i(k)$ from independence.
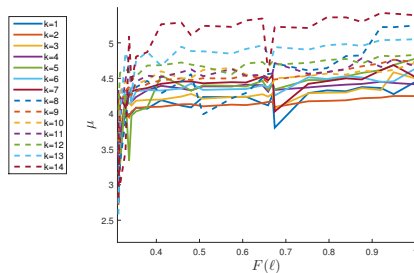- More mass along the diagonal and **a lot more** for matches of very high k and very high ℓ.
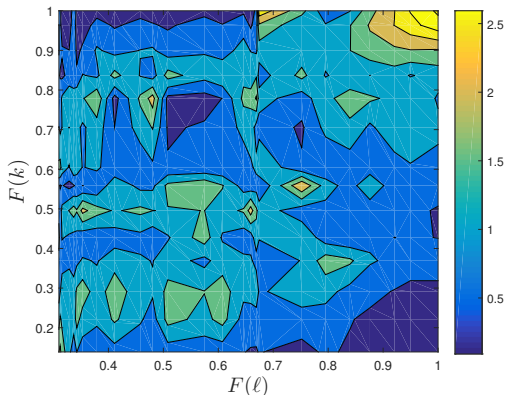
CONDITIONAL MEAN WAGES

Mean wage, $\mu_{k\ell}$

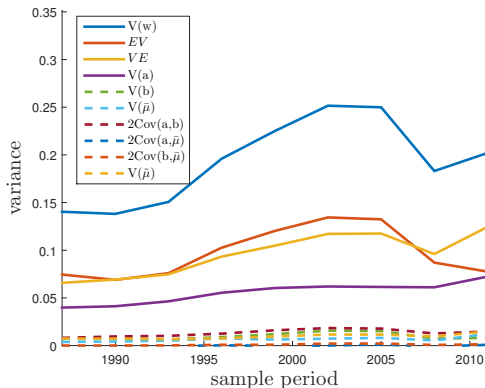Wages depend more on worker types. AKM's additive model not a bad approximation.

10-15 years of experience, tenure > 100 weeks

Unconditional variance decomposition

$$\text{Var}(w) = \underbrace{\text{Var}[\mu_{k\ell}(x)]}_{\text{between}} + \underbrace{\mathbb{E}[\sigma_{k\ell}(x)]}_{\text{idiosyncratic}}$$
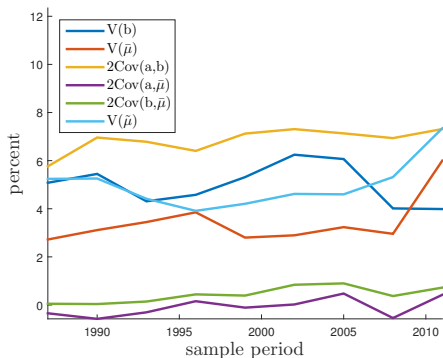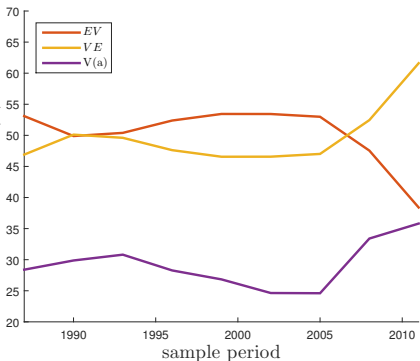
$$\text{Var}[\mu_{k\ell}(x)] = \text{Var}[\overline{\mu}(x)] + \text{Var}[a_k] + \text{Var}[b_\ell] + 2\text{Cov}[a_k, b_\ell]$$
$$+ 2\text{Cov}[\overline{\mu}(x), a_k] + 2\text{Cov}[\overline{\mu}(x), b_\ell] + \text{Var}[\widetilde{\mu}_{k\ell}(x)]$$

Unconditional variance decomposition

$$\text{Var}(w) = \underbrace{\text{Var}[\mu_{k\ell}(x)]}_{\text{between}} + \underbrace{\mathbb{E}[\sigma_{k\ell}(x)]}_{\text{idiosyncratic}}$$

$$\text{Var}[\mu_{k\ell}(x)] = \text{Var}[\bar{\mu}(x)] + \text{Var}[a_k] + \text{Var}[b_\ell] + 2\text{Cov}[a_k, b_\ell]$$
$$+ 2\text{Cov}[\bar{\mu}(x), a_k] + 2\text{Cov}[\bar{\mu}(x), b_\ell] + \text{Var}[\widetilde{\mu}_{k\ell}(x)]$$

Variance decomposition - Comparison with AKM

We find lower contributions of person and firm effects, bigger contribution of residuals, less increase in contribution of sorting.

		AKM	LPR			
		87-13	87-89	99-01	11-13	average
Residual	$E\sigma^2$	39.0	52.4	52.6	37.9	50.5
Person effect	Va	42.9	29.1	28.4	36.3	30.2
Firm effect	Vb	11.6	5.3	5.3	3.8	5.0
Cross effect	$2\mathrm{Cov}(a,b)$	3.3	5.6	7.3	7.3	6.9
Match effect	$V\widetilde{\mu}$		5.2	3.4	7.5	4.6
Observed heterogeneity	$V\overline{\mu}$	1.8	2.7	2.7	6.3	3.4
	$2\mathrm{Cov}(a,x)$	0.88	-0.21	-0.12	0.14	-0.10
	$2\mathrm{Cov}(b,x)$	0.52	-0.03	0.40	0.71	0.41
Sorting	$\mathrm{Corr}(a,b)$	7.4	22.5	30.0	31.0	28.9

Variance decomposition - Comparison with other studies

We find lower contributions of person and firm effects, bigger contribution of residuals, less increase in contribution of sorting.

	AKM											BLM	LPR
	FR	US1	AU	US2	IT	BR	DE		US3		DK	SW	DK
	76-87	84-93	90-97	90-99	81-97	95-05	85-91	02-09	80-86	07-13	87-13	02-04	average
$E\sigma^2$	15.8	9.3	5.0	9.3	15.1	7.0	7.7	5.0	20.3	14.9	39.0	25.2	50.5
Va	76.9	81.6	66.3	63.7	43.9	60.0	61.0	51.2	47.5	52.8	42.9	60.1	30.2
Vb	30.2	19.2	37.0	15.4	13.1	26.9	18.5	21.2	16.0	11.9	11.6	2.5	5.0
2Cov(a,b)	-27.2	-2.0	-22.4	0.62	2.1	3.2	2.3	16.4	1.6	7.1	3.3	12.2	6.9
V$\tilde{\mu}$				5.1			2.6	2.3					4.6
V$\overline{\mu}$	6.8	52.0	3.1	4.0	7.5	3.0	10.7	2.8	7.6	7.2	1.8		3.4
2Cov(a,x)	-3.1	-69.0	9.4	0.64	15.5		-2.6	0.70			0.88		-0.10
2Cov(b,x)	0.7	9.0	1.7	1.25	2.6		1.6	1.7			0.52		0.41
Corr(a,b)	-28.3	-2.5	-22.7	1.0	4.4	4.0	3.4	24.9	2.9	14.2	7.4	49.1	28.9

Abowd, Creecy, Krammarz (02), Gruetter, Lalive (09), Woodcock (09), Iranzo, Schivardi, Tosetti (08), Card, Heining, Kline (13), Song et al. (15), de Melo (17)

Variance decomposition - MC simulation (10-15 yrs exp, 2011-13)

AKM tends to overestimate person and firm effects, and underestimate residual variance and sorting.

	Truth	LPR	AKM	BS
Person effect	36.5	36.5	51.3	
Firm effect	6.2	6.7	10.6	
Sorting	8.4	8.3	1.8	
Residual	41.8	42.7	34.6	
Observed Heterogeneity	1.3	1.4	1.6	
Match effect	5.4	4.1		
Correlation	28.0	26.7	3.8	42.7

Based on the estimated model with $K = 14$ and $L = 24$, we simulate 1,000,000 workers and 100,000 firms. Panel length is 7 years.

ANATOMY OF SORTING

Initial match probability, $\pi_k m_{k\ell}^0$

Evidence of initial sorting
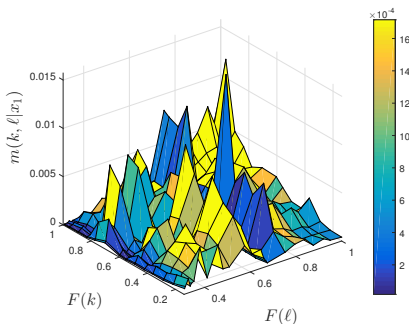
Observed, initial and equilibrium sorting

All sorting is obtained initially. Mobility maintains it.
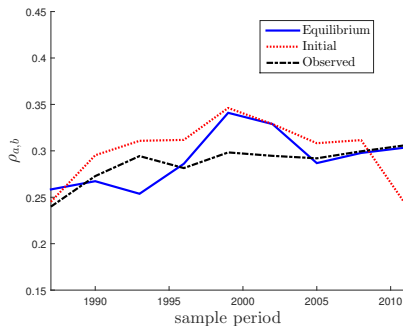
Correlation with $\mu_{k\ell}$

Stronger at low tenure than high tenure. Reallocation tends to take place subject to low tenure.

experience	$\sigma_{k\ell}$	$\delta_{k\ell}$	$\psi_{k\ell}$	$\gamma_{k\ell}$	λ_k	$\nu_{\ell'}$
	Low tenure					
<5 years	-0.41	-0.81	0.50	0.32	0.42	0.16
5-10 years	-0.28	-0.73	0.54	0.28	0.50	0.18
10-15 years	-0.19	-0.64	0.56	0.25	0.47	0.22
> 15 years	-0.16	-0.58	0.53	0.20	0.33	0.11
	High tenure					
<5 years	0.10	-0.28	0.28	0.05	0.56	0.39
5-10 years	0.13	-0.29	0.19	-0.12	0.61	0.39
10-15 years	0.15	-0.26	0.16	-0.10	0.58	0.42
> 15 years	0.14	-0.17	0.07	-0.17	0.35	0.41

σ: idiosyncratic sd; δ: layoff rate; ψ: employment finding rate; γ: preference for the job; λ : search intensity; ν: firm sampling rate

-> Details on parameters

Sorting through mobility: + 5 year $\mathbb{E}\left(F(\ell)'|F(\ell)\right)$

Sorting through mobility for young workers.
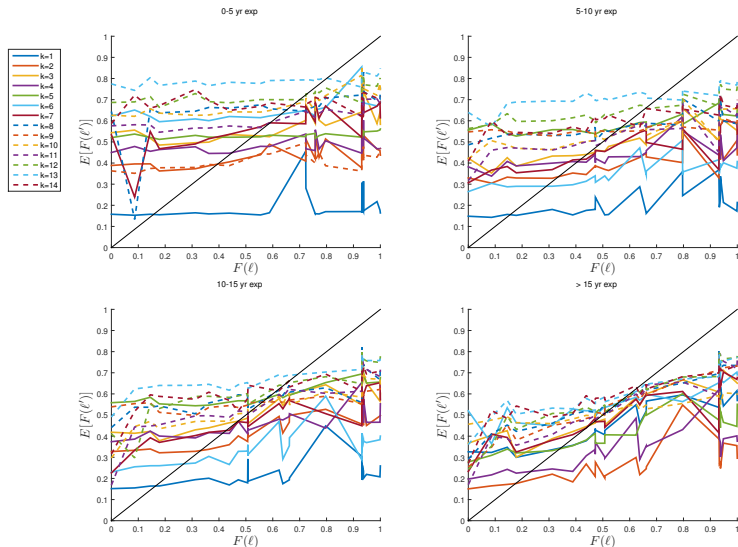
Synthetic Cohorts

- Create cohort starting with zero experience and zero tenure for a given time period.
- Initialize by the estimated $m_{k,\ell}(x)$. Simulate forward 30 years, holding calendar time fixed.
- Illustrates estimated sorting model interaction with tenure and experience.

Synthetic Cohorts - $\rho_{a,b}$ by cohort experience

Significant sorting from initial distribution. First 5 years of career strengthens sorting. Sorting subsequently declines as more experience is accumulated.
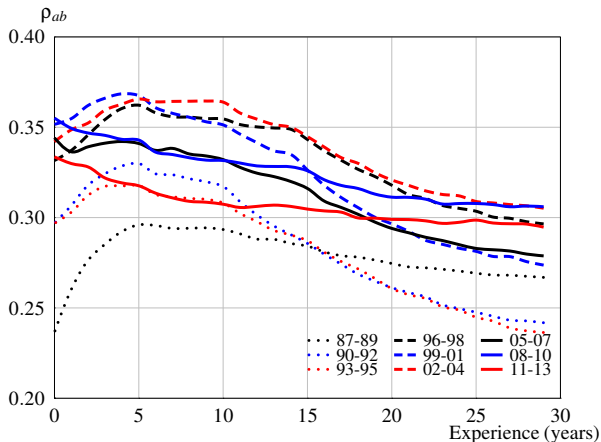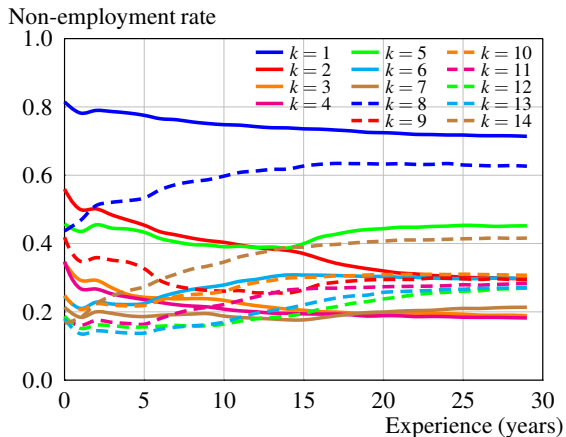
Synthetic Cohorts - Avg Firm effect $E[F(\ell)]$.

Higher type workers on average matched with higher firm effect firms. Lower type workers improve more on firm position early in career than higher type.
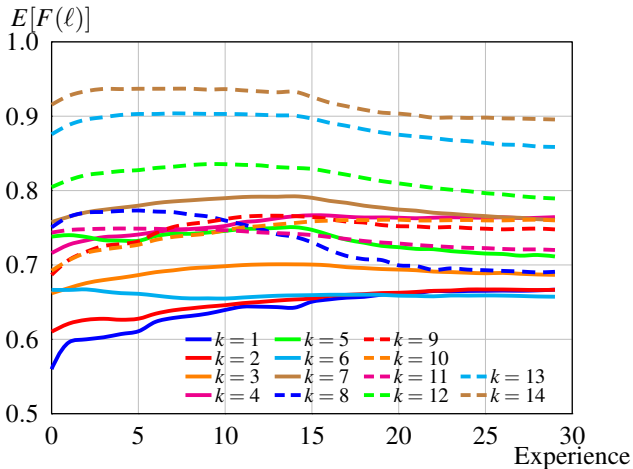
Synthetic Cohorts - Non-employment Rate.

Strong sorting on extensive margin. Low types have much higher non-employment rates.

Non-employment rate

Counterfactual Exercises - Unobserved types

Differences in re-employment risk, preferences for firms, and firm sampling probabilities are the key drivers of this mechanism of resistance to disorder.

	Average $\rho_{a,b}$	% change
Benchmark	0.28	
Counterfactual		
No k variation in $\gamma_{k\ell}$	0.10	-63.5
No k variation in $\psi_{k\ell}$	0.18	-33.6
No k variation in λ_k	0.27	-2.9
No k variation in $\delta_{k\ell}$	0.28	0.4
No ℓ variation in ν_ℓ	0.12	-57.6
No U-shock, zero $\delta_{k\ell}$	0.24	-10.5
No E-E transition, zero λ_k	0.17	-39.5

CONCLUSION

Concluding remarks

- Following discrete mixture approach in BLM, we present a CEM+MM algorithm for the flexible estimation of wage and mobility parameters.
  - Fast estimation of nonlinear mobility parameters from MM algorithm in M-step.
  - C-step improves performance of estimator.
- Main Findings
  - Cross-sectional wage distributions strongly depend on worker type.
  - Sizable degree of sorting over time, essentially stable
  - AKM tends to overestimate person and firm effects, and underestimate residual variance and sorting.
  - Most sorting is obtained when young. Early career mobility strengthens sorting. Subsequent mobility tends to undo it.
  - Differences in re-employment risk, mobility out of non-employment, and preferences for firms are the key drivers to sustain sorting.

DETAILS ON CEM ALGORITHM

Likelihood given firm classification

- Let D_{it} indicate an employer change between t and $t+1$ by,

$$D_{it} = \begin{cases} 1 & \text{if } j_{i,t+1} \neq j_{it} \\ 0 & \text{if } j_{i,t+1} = j_{it}. \end{cases}$$

- For a given firm classification $F = (\ell_1, ..., \ell_J)$, let the share of type-$\ell$ firms be

$$q_\ell(F) = \frac{\#\{j : \ell_j = \ell\}}{J}$$

- Let ℓ_{it} be the type of worker i's employer in period t.

Likelihood given firm classification

- The likelihood of a worker history $(w_{it}, j_{it}, x_{it})_{t=1}^{T_i}$ given observed worker types (z_i) and firm types $F = (\ell_1, ..., \ell_J)$ is,

$$L_i(\beta, F) = \sum_{k=1}^{K} L_i(k; \beta, F),$$

where $\beta = (\mu, \sigma, \lambda, \nu, \gamma, \pi, m^0)$ and

$$L_i(k; \beta, F) = \frac{\pi_k(z_i) \, m_{k,\ell_{i1}}^0(x_{i1})}{q_{\ell_{i1}}(F)} \prod_{t=1}^{T} f_{k\ell_{it}}(w_{it}|x_{it})$$
$$\times \prod_{t=1}^{T-1} M_{k\ell_{it}\neg}(x_{it})^{1-D_{it}} \left(\frac{M_{k\ell_{it}\ell_{i,t+1}}(x_{it})}{q_{\ell_{i,t+1}}(F)} \right)^{D_{it}},$$

is the likelihood of worker i's observed history and that i is type k.
  - Note: $\frac{1}{q_{\ell_{i1}}(F)} \propto$ probability of j_{i1} given $\ell_{j_{i1}} = \ell_{i1}$ (uniform sampling)

- We use the EM algorithm for estimation $\widehat{\beta}$ of β given a firm classification F.

E step

Posterior probability of worker type:

$$p_i(k; \beta^{(m)}, F) = \frac{L_i(k; \beta^{(m)}, F)}{\sum_{k=1}^{K} L_i(k; \beta^{(m)}, F)}.$$

M step

- Wage distributions (OLS):

$$(\mu, \sigma)^{(m+1)} = \arg \max_f \sum_{i,k} p_i(k; \beta^{(m)}, F) \left[\sum_{t=1}^{T} \ln f_{k\ell_{it}}(w_{it}|x_{it}) \right]$$

- Transition probabilities:

$$(\lambda, \nu, \gamma)^{(m+1)} = \arg \max_M \sum_{i,k} p_i(k; \beta^{(m)}, F)$$

$$\times \left(\sum_t \left[(1 - D_{it}) \ln M_{k\ell_{it}\neg}(x_{it}) + D_{it} \ln M_{k\ell_{it}\ell_{i,t+1}}(x_{it}) \right] \right)$$

Non-linear estimation. Here, we adapt Hunter's (2004) MM-estimator for the Bradley-Terry model.

- Initial probabilities (simple frequencies):

$$\pi_k^{(m+1)}(z) = \frac{\sum_{i:z_i=z} p_i(k; \beta^{(m)}, F)}{\#\{i : z_i = z\}}, \ m_{k\ell_1}^{(m+1)}(x_1) = \frac{\sum_{ix_{i1}=x_1, \ell_{i1}=\ell_1} p_i(k; \beta^{(m)}, F)}{\sum_{i:x_{i1}=x_1} p_i(k; \beta^{(m)}, F)}$$

Firm classification update

- Order firms j by decreasing size.
- Let $\widehat{\beta}^{(s)}$ be EM-estimator of β given firm classification $F^{(s)}$.
- Given $\widehat{\beta}^{(s)}, F^{(s)}$, we update $\ell_j^{(s)}$ iteratively as

$$\ell_j^{(s+1)} = \arg\max_{\ell_j} \sum_{i,k} p_i(k|\widehat{\beta}^{(s)}, F^{(s)}) \ln L_i(k; \widehat{\beta}^{(s)}, F_{j-}^{(s+1)}, \ell_j, F_{j+}^{(s)})$$

  where $F_{j-}^{(s+1)} = (\ell_1^{(s+1)}, ..., \ell_{j-1}^{(s+1)})$ and $F_{j+}^{(s)} = (\ell_{j+1}^{(s)}, ..., \ell_J^{(s)})$
- Guarantees likelihood improvement in each iteration.
- Monte Carlo simulations show that our reclassification algorithm improves pre-classification by k-means algorithm.
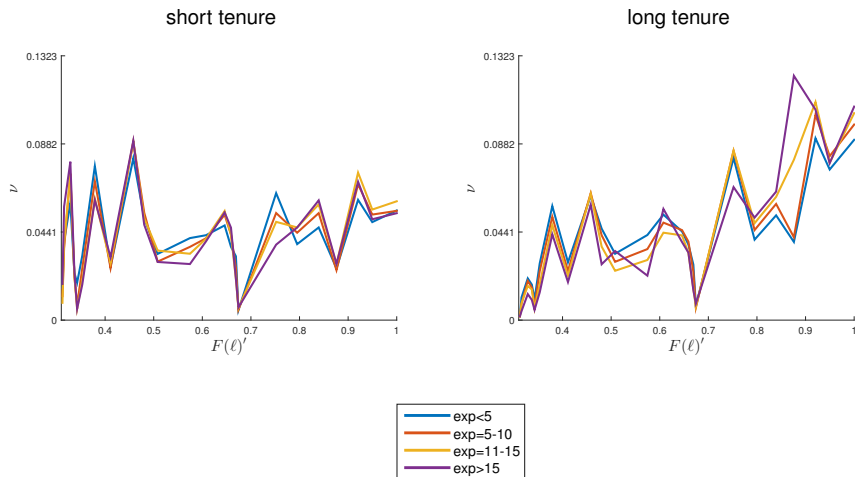
back

PARAMETERS

Search intensity λ_k

Weakly increasing with k. Tenure and experience effects.
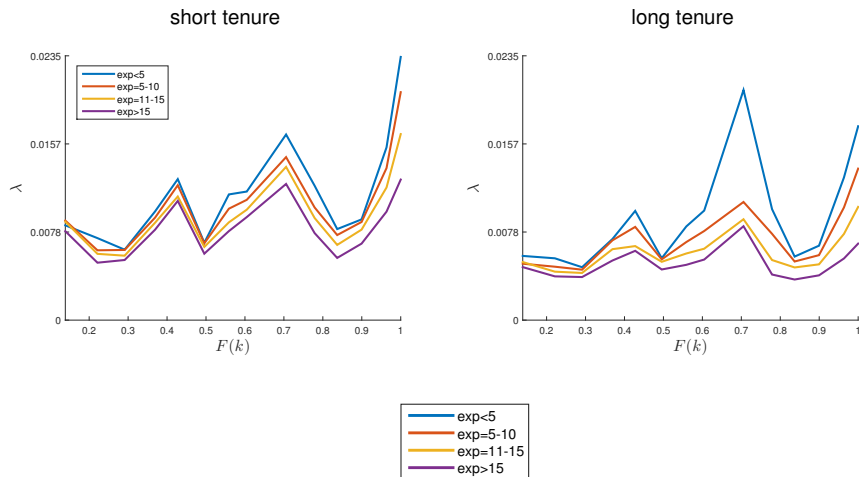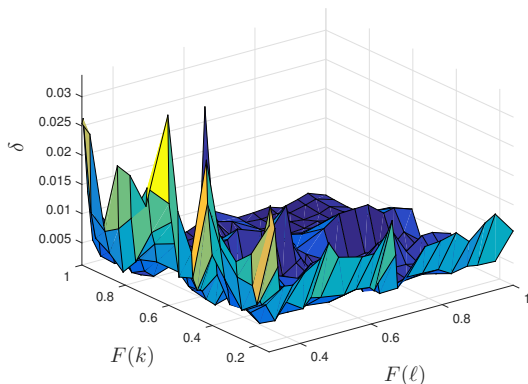
Firm sampling proba $\nu_{\ell'}$

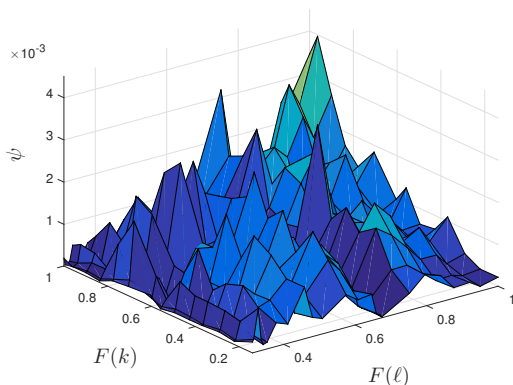Increases with ℓ'. Tenure effect.

Layoff rate, $\delta_{k\ell}$

High k, ℓ protected from layoff

10-15 years of experience, short tenure

Job finding rate for unemployed, $\psi_{k\ell}$

High k, ℓ are re-employed faster

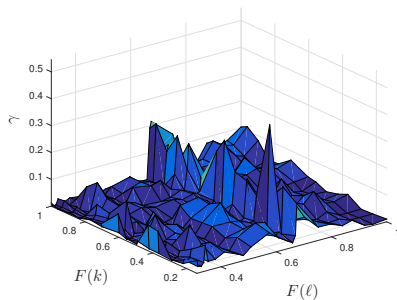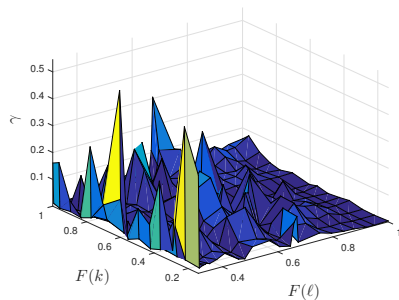# Preference for the job, $\gamma_{k\ell}$

Weakly increasing with bumps. Tenure matters.

10-15 years of experience

short tenure

long tenure

back