The Anatomy of Sorting

Evidence from Danish Data

Rasmus Lentz Suphanit Piyapromdee Jean-Marc Robin

DSE 2024 – August 5, 2024

Introduction

Mobility Determinants

- Are workers and firms sorted?
- If so,
  - in which way?
  - by which mobility mechanisms?

What is sorting?

- Workers and firms are sorted if in a given match, knowing one side's type is informative about the type of the other.
  - Measure by mutual information measure.

What is sorting?

- Workers and firms are sorted if in a given match, knowing one side's type is informative about the type of the other.
  - Measure by mutual information measure.
- If firm and worker types can be assigned cardinal labels, sorting can be measured by for example covariance or stochastic dominance.
  - Wage sorting is an example.
  - Abowd, Kramarz, and Margolis (1999), Bonhomme, Lamadon, and Manresa (2019), Kline, Saggio, and Sølvsten (2020).

What we do

- Identify finite mixture model of wage and employment dynamics where sorting happens through 4 channels:
  - Job preferences (choice).
  - Market segmentation (chance).
  - Layoffs.
  - Unemployed job finding.

What we do

- Identify finite mixture model of wage and employment dynamics where sorting happens through 4 channels:
  - Job preferences (choice).
  - Market segmentation (chance).
  - Layoffs.
  - Unemployed job finding.
- Broaden study of sorting to include non-wage determinants.

What we do

- Identify finite mixture model of wage and employment dynamics where sorting happens through 4 channels:
  - Job preferences (choice).
  - Market segmentation (chance).
  - Layoffs.
  - Unemployed job finding.
- Broaden study of sorting to include non-wage determinants.
- Builds on Bonhomme, Lamadon, and Manresa (2019) and Abowd, Mckinney, and Schmutte (2018). Extend BLM by,
  - Classification EM algorithm. Ensures that firm classification uses observed mobility and firms' labor force worker type composition in addition to wage information (important in case sorting is not on wages).
  - Extend EM algorithm to quickly and efficiently identify mobility model.
  - Allow rich observed heterogeneity to co-determine mobility and wages.

Key Findings

- Estimate on Danish matched employer-employee data, 1989-2013.

Key Findings

- Estimate on Danish matched employer-employee data, 1989-2013.
- Choice (preferences):
  - Workers perceive clearly defined “job type ladders”.
  - Preferences intensify with tenure and experience.
  - Preferences intensify in worker wage type.
  - Significant pecuniary component to job preferences (in particular NPV of future earnings). Weakens with age.

Key Findings

- Estimate on Danish matched employer-employee data, 1989-2013.
- Choice (preferences):
  - Workers perceive clearly defined “job type ladders”.
  - Preferences intensify with tenure and experience.
  - Preferences intensify in worker wage type.
  - Significant pecuniary component to job preferences (in particular NPV of future earnings). Weakens with age.
- Type sorting (MI) strengthens as workers age and tenure rises, but in line with decreased preference emphasis on wages, wage sorting does not.

Key Findings

- Estimate on Danish matched employer-employee data, 1989-2013.
- Choice (preferences):
  - Workers perceive clearly defined “job type ladders”.
  - Preferences intensify with tenure and experience.
  - Preferences intensify in worker wage type.
  - Significant pecuniary component to job preferences (in particular NPV of future earnings). Weakens with age.
- Type sorting (MI) strengthens as workers age and tenure rises, but in line with decreased preference emphasis on wages, wage sorting does not.
- Layoff channel contributes less than other three channels to sorting, but all channels are substantial.

Key Findings

- Estimate on Danish matched employer-employee data, 1989-2013.
- Choice (preferences):
  - Workers perceive clearly defined “job type ladders”.
  - Preferences intensify with tenure and experience.
  - Preferences intensify in worker wage type.
  - Significant pecuniary component to job preferences (in particular NPV of future earnings). Weakens with age.
- Type sorting (MI) strengthens as workers age and tenure rises, but in line with decreased preference emphasis on wages, wage sorting does not.
- Layoff channel contributes less than other three channels to sorting, but all channels are substantial.
- Market segmentation is key driver of sorting during early career, whereas job preferences dictate matching when old.

Model

Matches and agent heterogeneity

- A job is a match between a worker and a firm.
- A worker is characterized by (k, x),
  - latent type $k \in \{1, \dots, K\}$.
  - observable, time varying characteristics, x.
- A firm is characterized by latent type $\ell \in \{1, \dots, L\}$.
- At any given time, a worker can be matched with at most one firm or be non-employed.
- A firm can be matched with many workers.

Match Wages

- Match wages are AR(1) (within-match) and log-normally distributed.
- Specifically, the initial log wage, w, is distributed according to,

$$f^0_{k\ell}(w|x) = \frac{1}{\omega_{k\ell}(x)}\varphi\left(\frac{w-\mu_{k\ell}(x)}{\omega_{k\ell}(x)}\right),$$

and the one-period-forward log wage, w', according to,

$$f_{k\ell}(w'|x', x, w) = \frac{1}{\sigma_{k\ell}(x')}\varphi\left(\frac{w'-\mu_{k\ell}(x')-\rho\left[w-\mu_{k\ell}(x)\right]}{\sigma_{k\ell}(x')}\right).$$

  - $\mu_{k\ell}(x)$ is a k-worker's average log-wage when matched with an ℓ-firm.
  - $\left(\omega_{k\ell}(x), \sigma_{k\ell}(x)\right)$ are the standard deviations of the noise innovations.
  - ρ is the AR coefficient.
  - $\varphi(\cdot)$ is the Gaussian kernel.

Job Mobility

- Employed k-worker meets outside type-ℓ' job opportunity with probability, $\lambda_{k\ell'}(x)$.

Job Mobility

- Employed k-worker meets outside type-ℓ' job opportunity with probability, $\lambda_{k\ell'}(x)$.
- The worker accepts the outside offer with probability,

$$P_{k\ell\ell'}(x) = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)},$$

where $\gamma_{k\ell}(x) = \exp\big(V_{k\ell}(x)/\nu_{kx}\big)$ reflects the worker's valuation of a match with an ℓ-firm.

- *Random utility interpretation:* At point of choice, each match is subject to iid Gumbel distributed value addition. ν_{kx} variance parameter.
- *Random mobility cost interpretation:* At point of choice, mobility cost realization drawn from logit distribution. ν_{kx} variance parameter.
- Sorkin (2018) proposes a restricted version.
- Logit choice in search also in Lentz et al. (2023) and Arcidiacono, et al. (2023)
- Lamadon et al. (2024) a very promising example of fuller structural interpretation that fleshes out amenity implications. Uses this setup as first step. See Jeremy Lise's presentation at conference.

Job Mobility

- Employed k-worker meets outside type-ℓ' job opportunity with probability, $\lambda_{k\ell'}(x)$.
- The worker accepts the outside offer with probability,

$$P_{k\ell\ell'}(x) = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)},$$

  where $\gamma_{k\ell}(x) = \exp\big(V_{k\ell}(x)/\nu_{kx}\big)$ reflects the worker's valuation of a match with an ℓ-firm.
  - *Random utility interpretation:* At point of choice, each match is subject to iid Gumbel distributed value addition. ν_{kx} variance parameter.
  - *Random mobility cost interpretation:* At point of choice, mobility cost realization drawn from logit distribution. ν_{kx} variance parameter.
  - Sorkin (2018) proposes a restricted version.
  - Logit choice in search also in Lentz et al. (2023) and Arcidiacono, et al. (2023)
  - Lamadon et al. (2024) a very promising example of fuller structural interpretation that fleshes out amenity implications. Uses this setup as first step. See Jeremy Lise's presentation at conference.
- $M_{k\ell\ell'}(x) = \lambda_{k\ell'}(x)P_{k\ell\ell'}(x)$ for $\ell, \ell' > 0$.

Job Mobility

- Employed k-worker meets outside type-ℓ' job opportunity with probability, $\lambda_{k\ell'}(x)$.
- The worker accepts the outside offer with probability,

$$P_{k\ell\ell'}(x) = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)},$$

where $\gamma_{k\ell}(x) = \exp\big(V_{k\ell}(x)/\nu_{kx}\big)$ reflects the worker's valuation of a match with an ℓ-firm.
  - *Random utility interpretation:* At point of choice, each match is subject to iid Gumbel distributed value addition. ν_{kx} variance parameter.
  - *Random mobility cost interpretation:* At point of choice, mobility cost realization drawn from logit distribution. ν_{kx} variance parameter.
  - Sorkin (2018) proposes a restricted version.
  - Logit choice in search also in Lentz et al. (2023) and Arcidiacono, et al. (2023)
  - Lamadon et al. (2024) a very promising example of fuller structural interpretation that fleshes out amenity implications. Uses this setup as first step. See Jeremy Lise's presentation at conference.
- $M_{k\ell\ell'}(x) = \lambda_{k\ell'}(x)P_{k\ell\ell'}(x)$ for $\ell, \ell' > 0$.
- Employed k-worker is laid off into non-employment with probability $\delta_{k\ell}(x) = M_{k\ell 0}(x)$.
- Non-employed k-worker finds a job with an ℓ-firm with probability $\varphi_{k\ell}(x) = M_{k0\ell}(x)$.

Data and Estimation

DATA

- Danish register data, 1989-2013.
- Divided in five 5-year periods: 89-93, 94-98, 99-03, 04-08, 09-13.
- Data on weekly wages, worker and employer IDs.
- Observable worker characteristics: education, gender, potential experience, tenure in the job. Categorize into:
  - 8 time varying worker characteristics groups (experience×tenure), x.
  - 6 time invariant worker characteristics groups (education×gender), z^w.
- Observable firm characteristics: sector (public, private, mixed), industry and age, z^f.

Worker and firm classification

- Adopt the finite mixture approach of Bonhomme, Lamadon and Manresa (2019)
- Worker latent types $k = 1, \ldots, K$; firm latent types $\ell = 1 \ldots, L$.
- In estimation firm classification denoted by $F = (\ell_1, \ldots, \ell_J)$ treated like a fixed effect.

Initial state

- Initial distribution,
$$P^0\big(k,\ell,x,z^w,z^f\big) = m^0_{k\ell}(x)\,\pi\big(z^w \mid k\big)\,\pi\big(z^f \mid \ell\big),$$
where $x = (\text{ten}, \text{exp})$, $z^w = (\text{edu}, \text{sex})$, $z^f = (\text{sector}, \text{industry})$
- Latent types are sufficient statistics for time invariant heterogeneity.
- Observed time invariant heterogeneity aid in classification.

LIKELIHOOD GIVEN FIRM CLASSIFICATION

- The complete likelihood of worker i's history and that s/he is type k is

$$\mathcal{L}_i(k;\beta,F) = \frac{m^0_{k\ell_{i1}}(x)\pi(z^w|k)\pi(z^f|\ell_{i1})}{\#(\ell_{i1}|F)} \prod_{t=1}^{T-1} M_{k\ell_{it}\neg}(x_{it})^{1-D_{it}} \left(\frac{M_{k\ell_{it}\ell_{i,t+1}}(x_{it})}{\#(\ell_{i,t+1}|F)}\right)^{D_{it}} \times$$
$$\prod_{t=1}^{T} f^0_{k\ell_{it}}(w_{it}|x_{it})^{D_{i,t-1}} f_{k\ell_{it}}(w_{it}|x_{it})^{1-D_{i,t-1}},$$

  where
  - $D_{it} = 1$ if employment transition (E-U, U-E and E-E). $D_{i0} = 1$.
  - $F = (\ell_1, \dots, \ell_J)$ is firm classification.
  - $\ell_{it} = \ell_{j_{it}}$, where j_{it} is firm matched to worker i at time t. $\ell_{it} = 0$ if unmatched.
  - $\beta = (\mu, \sigma, \omega, \rho, \lambda, \gamma, \varphi, \delta, \pi, m^0)$.

Likelihood given firm classification

- The complete likelihood of worker i's history and that s/he is type k is

$$\mathcal{L}_i(k;\beta,F)=\frac{m^0_{k\ell_{i1}}(x)\pi(z^w|k)\pi(z^f|\ell_{i1})}{\#(\ell_{i1}|F)}\prod_{t=1}^{T-1}M_{k\ell_{it}\neg}(x_{it})^{1-D_{it}}\left(\frac{M_{k\ell_{it}\ell_{i,t+1}}(x_{it})}{\#(\ell_{i,t+1}|F)}\right)^{D_{it}}\times$$

$$\prod_{t=1}^{T}f^0_{k\ell_{it}}(w_{it}|x_{it})^{D_{i,t-1}}f_{k\ell_{it}}(w_{it}|x_{it})^{1-D_{i,t-1}},$$

  where
  - $D_{it}=1$ if employment transition (E-U, U-E and E-E). $D_{i0}=1$.
  - $F=(\ell_1,\dots,\ell_J)$ is firm classification.
  - $\ell_{it}=\ell_{j_{it}}$, where j_{it} is firm matched to worker i at time t. $\ell_{it}=0$ if unmatched.
  - $\beta=(\mu,\sigma,\omega,\rho,\lambda,\gamma,\varphi,\delta,\pi,m^0)$.
- The marginal likelihood is $\mathcal{L}_i(\beta,F)=\sum_{k=1}^K\mathcal{L}_i(k;\beta,F)$.

Likelihood given firm classification

- The complete likelihood of worker i's history and that s/he is type k is

$$\mathcal{L}_i(k;\beta,F) = \frac{m^0_{k\ell_{i1}}(x)\pi(z^w|k)\pi(z^f|\ell_{i1})}{\#(\ell_{i1}|F)} \prod_{t=1}^{T-1} M_{k\ell_{it}\neg}(x_{it})^{1-D_{it}} \left(\frac{M_{k\ell_{it}\ell_{i,t+1}}(x_{it})}{\#(\ell_{i,t+1}|F)}\right)^{D_{it}} \times$$
$$\prod_{t=1}^{T} f^0_{k\ell_{it}}(w_{it}|x_{it})^{D_{i,t-1}} f_{k\ell_{it}}(w_{it}|x_{it})^{1-D_{i,t-1}},$$

  where
  - $D_{it} = 1$ if employment transition (E-U, U-E and E-E). $D_{i0} = 1$.
  - $F = (\ell_1, \ldots, \ell_J)$ is firm classification.
  - $\ell_{it} = \ell_{j_{it}}$, where j_{it} is firm matched to worker i at time t. $\ell_{it} = 0$ if unmatched.
  - $\beta = (\mu, \sigma, \omega, \rho, \lambda, \gamma, \varphi, \delta, \pi, m^0)$.
- The marginal likelihood is $\mathcal{L}_i(\beta, F) = \sum_{k=1}^{K} \mathcal{L}_i(k;\beta,F)$.
- You may notice that were data complete, maximization of log of complete likelihoods would involve simple first order conditions in β (almost - restricted mobility model a complication).
- But, sadly, data are not complete.

A (V)EM ASIDE

- Let X be observed data. $Z = (Z^w, Z^f)$ is missing data. In our setup, it is the latent types.
- Denote by $\ln \mathcal{L}(X)$ the log-likelihood of the data given parameterization β.
- For any distribution $R(Z)$ (and using $\mathcal{L}$ also for probabilities, when appropriate).

$$\begin{aligned}
\ln \mathcal{L}(X) = \sum_Z R(Z) \ln \mathcal{L}(X) &= \sum_Z R(Z) \ln \left(\frac{\mathcal{L}(X, Z)}{\mathcal{L}(Z \mid X)} \right) \\
&= \sum_Z R(Z) \ln \mathcal{L}(X, Z) - \sum_Z R(Z) \ln \mathcal{L}(Z \mid X) \\
&\geq \sum_Z R(Z) \ln \mathcal{L}(X, Z) - \sum_Z R(Z) \ln R(Z),
\end{aligned}$$

  where inequality obtains from Gibbs, $-\sum_Z R(Z) \ln R(Z) \leq -\sum_Z R(Z) \ln \mathcal{L}(Z \mid X)$.
- If $\mathcal{L}(Z \mid X)$ is tractable, then EM algorithm is available to maximize likelihood.
- BLM (2019) insight: $\mathcal{L}(Z^w \mid X, Z^f)$ is tractable. Get Z^f in pre-step k-means. Green light!
- LPR (2023): Extend to CEM algorithm. Iteratively update Z^f to improve on likelihood.
- Hong, Lentz, Robin (2024): Adopt VEM to obtain Z^f - Choose tractable $R(Z)$ to minimize distance to $\ln \mathcal{L}(Z \mid X)$.

EM is MM

- MM: Minorization-Maximization or Majorization-Minimization.
- Let β be a parameterization of the likelihood.
- Define
$$H^{(m)}(X,\beta) = \sum_Z R^{(m)}(Z) \ln \mathcal{L}(X,Z,\beta) - \sum_Z R^{(m)}(Z) \ln R^{(m)}(Z),$$
where $R^{(m)}(Z) = \mathcal{L}(Z \mid X, \beta^{(m)})$
- By previous slide,
$$\ln \mathcal{L}(X,\beta) \geq H^{(m)}(X,\beta),$$
with equality at $\beta = \beta^{(m)}$.
- That is, $H^{(m)}(X,\beta)$ is a minorization of $\ln \mathcal{L}(X,\beta)$ in the point $\beta^{(m)}$.
- Implication: Any $\beta : H^{(m)}(X,\beta) > H^{(m)}(X,\beta^{(m)}) \Rightarrow \mathcal{L}(X,\beta) > \mathcal{L}(X,\beta^{(m)})$.
- EM:
$$\beta^{(m+1)} = \arg\max_\beta H^{(m)}(X,\beta) = \arg\max_\beta \sum_Z R^{(m)}(Z) \ln \mathcal{L}(X,Z,\beta).$$

CEM Estimation

- E step: given a firm type vector $F^{(s)}$, define posterior worker-type probability,

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

- M step: take $p_i(k;\beta^{(m)},F^{(s)})$ as given, compute $\beta^{(m+1)}$ to maximize (modified) expected log-likelihood. MM-algorithm.
- C step: Given $\widehat{\beta}^{(s)}, F^{(s)}$, we update each $\ell_j^{(s)}$ as

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

- Iterate on EM using $F^{(s+1)}$ until convergence then repeat C step until CEM converges.

Number of groups

- We estimate L first using k-means, seeking to maximize the between-group/within-group variance ratio.
- We thus pick $L = 14$ in periods 1-2 and $L = 22$ for periods 3-5.
- We fix $K = 24$ which is about the maximum given memory constraints.

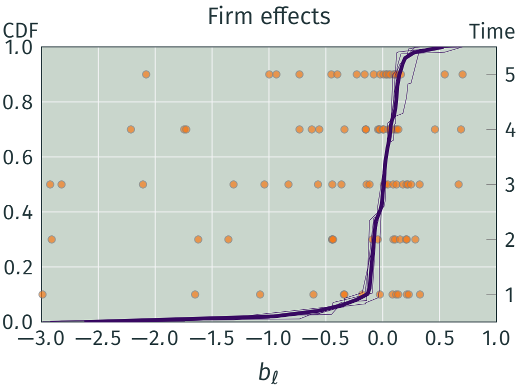
- To evaluate the extent of wage sorting, we use the linear projection,

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

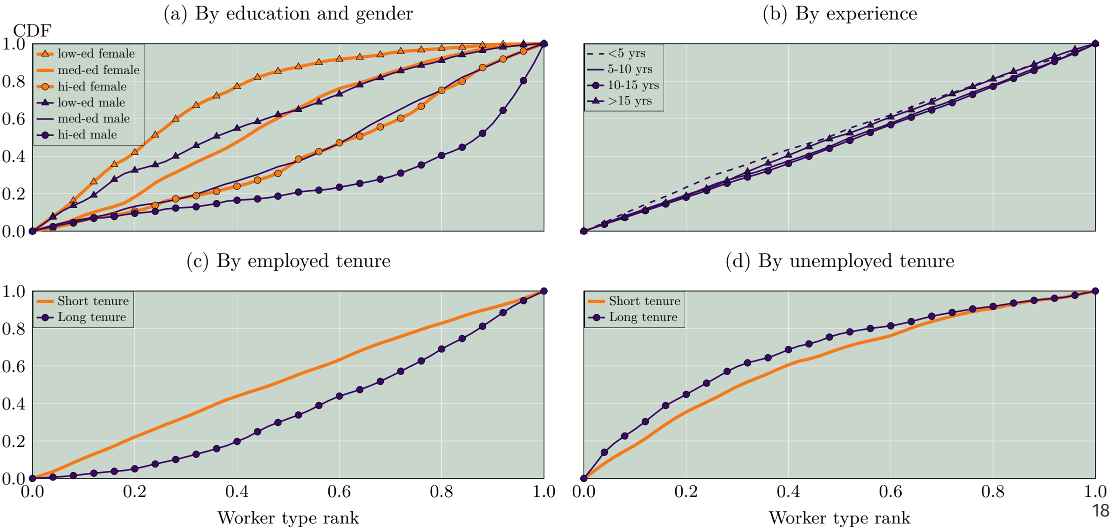
  where $\overline{\mu}(x)$ contains tenure-experience interactions.
- Order worker types by a_k (k's wage label).
- Order firm types by b_ℓ (ℓ's wage label).

Wage labels

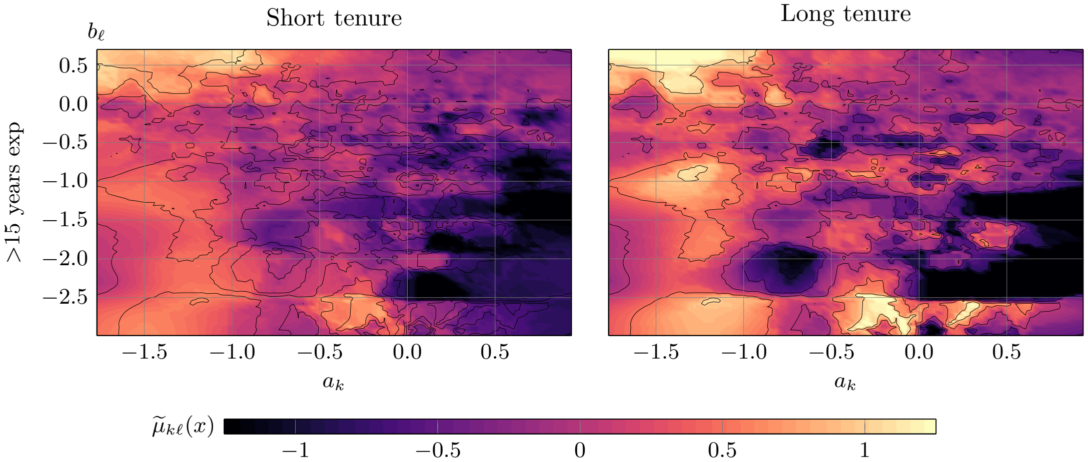

Worker effects
CDF
Time
1.0
0.8
0.6
0.4
0.2
0.0
5
4
3
2
1
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
a_k


Firm effects
CDF
Time
1.0
0.8
0.6
0.4
0.2
0.0
5
4
3
2
1
−3.0 −2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
b_ℓ

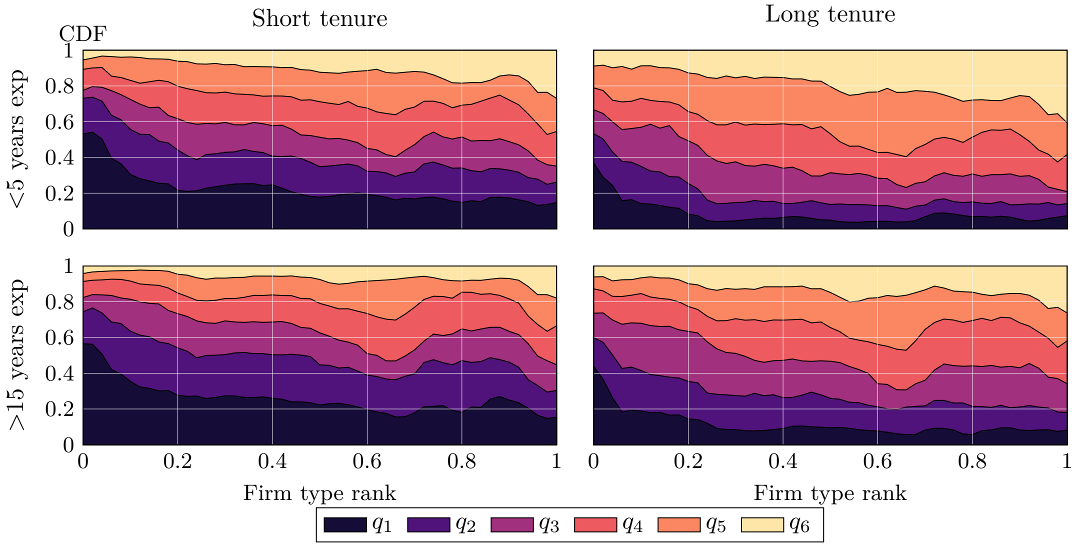
Latent wage types by observed worker characteristics



Wage residuals



Matching by wage label ranks



Log-wage variance decomposition

	89-93	94-98	99-03	04-08	09-13
residual	42.7	46.4	49.0	50.8	39.6
worker effect, a_k	23.6	22.7	19.3	22.8	25.8
match effect, $\widetilde{\mu}_{k\ell}(x)$	15.6	14.2	13.1	11.5	10.4
firm effect, b_ℓ	8.3	7.0	8.8	5.0	5.4
sorting effect, $2\,\mathrm{Cov}(a_k, b_\ell)$	5.4	5.0	6.1	5.3	6.8
tenure and experience, x	4.4	4.8	4.2	4.6	11.9
$\overline{\mu}(x)$	1.3	2.4	1.5	2.5	4.4
$2\,\mathrm{Cov}(a_k, \overline{\mu}(x))$	2.6	2.0	1.9	1.7	5.9
$2\,\mathrm{Cov}(\overline{\mu}(x), b_\ell)$	0.5	0.4	0.3	0.4	1.6
total between	57.3	53.6	51.0	49.2	60.4
wage correlation, $\mathrm{Cor}(a_k, b_\ell)$	0.19	0.20	0.23	0.25	0.29

Note: Percent of total variance.

Parameter Variance Decompositions - Avg over 5-yr windows

	$\mu_{k\ell}(x)$	$\delta_{k\ell}(x)$	$\lambda_{k\ell}(x)$	$\psi_{k\ell}(x)$	$\gamma_{k\ell}(x)$
Coefficient of variation	0.08	1.43	1.36	1.08	0.96
Variance decomposition (%)					
worker effect, a_k^y	42.02	45.93	36.57	16.47	13.51
match effect, $\widetilde{y}_{k\ell}(x)$	23.98	29.14	40.42	43.32	52.60
firm effect, b_ℓ^y	12.76	6.51	12.24	22.20	23.56
sorting effect, $2\,\mathrm{Cov}(a_k, b_\ell)$	10.57	4.09	−2.27	1.21	2.64
tenure and experience, x	10.67	14.33	13.05	16.80	7.69
$\overline{y}(x)$	4.42	6.11	6.32	18.95	2.99
$2\,\mathrm{Cov}(a_k, \overline{y}(x))$	5.10	7.02	7.78	2.32	2.84
$2\,\mathrm{Cov}(b_\ell, \overline{y}(x))$	1.15	1.20	−1.06	−4.47	1.86
$\mathrm{Cor}(a_k, b_\ell)$	0.23	0.11	−0.04	0.05	0.10

Measured job preferences

Choice and Chance

- E-E flows identify chance relative to choice

$$M_{k\ell\ell'}(x) = \underbrace{\lambda_{k\ell'}(x)}_{\text{chance}} \times \underbrace{\frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)}}_{\text{choice}}.$$

- In discrete choice w/ type 1 idiosyncratic taste shocks, $\ln\big(\gamma_{k,\ell}(x)\big) = V_{k\ell}(x)$.
- Measure preference intensity as Kullback-Leibler divergence from uniformity,

$$d_{KL}(\gamma) = \sum_{\ell=1}^{L} \gamma_{k\ell}(x) \ln\left(\frac{\gamma_{k\ell}(x)}{1/L}\right).$$

- ”Intensity” used as antonym for ”indifference”.

Preference intensity, $d_{KL}(\gamma)$

Experience (yrs):		Short tenure 0-5	5-10	10-15	15+	Long Tenure 0-5	5-10	10-15	15+
	Low	0.38	0.47	0.60	0.63	0.44	0.62	0.99	0.97
1989-93	Med	0.75	0.67	0.73	0.76	1.05	1.16	0.83	1.00
	High	1.03	0.89	0.91	1.12	1.06	0.99	1.01	1.17
	Low	0.37	0.34	0.48	0.51	0.84	0.65	0.85	0.77
1994-98	Med	0.63	0.59	0.65	0.85	0.81	0.67	0.73	0.92
	High	0.69	0.74	0.77	0.91	0.80	0.68	0.79	0.88
	Low	0.52	0.59	0.62	0.65	1.10	0.75	0.82	0.79
1999-03	Med	0.76	0.64	0.70	0.76	0.95	0.63	0.69	0.83
	High	0.84	0.86	0.95	1.09	0.94	0.95	0.88	1.04
	Low	0.51	0.56	0.55	0.62	1.18	0.98	0.87	0.70
2004-08	Med	0.81	0.79	0.92	0.95	0.98	0.94	1.09	0.95
	High	0.79	0.77	0.90	0.95	0.90	0.80	0.94	0.96
	Low	0.73	0.53	0.60	0.70	0.84	1.08	0.95	1.00
2009-13	Med	0.82	0.76	0.75	0.83	0.87	0.98	1.01	1.11
	High	0.99	0.93	0.98	1.01	1.06	1.09	1.14	1.18

- Intensity increasing in experience and tenure.
- Intensity increasing in worker’s wage type.

Mobility model component correlations

	Short tenure				Long tenure			
Experience:	0-5	5-10	10-15	15+	0-5	5-10	10-15	15+
$\gamma_{k\ell}(x), \mu_{k\ell}(x)$	0.41	0.43	0.42	0.40	0.22	0.25	0.20	0.20
$\delta_{k\ell}(x), \mu_{k\ell}(x)$	-0.25	-0.31	-0.32	-0.27	-0.12	-0.22	-0.23	-0.14
$\lambda_{k\ell}(x), \mu_{k\ell}(x)$	0.13	0.13	0.12	0.10	0.15	0.20	0.22	0.21
$\psi_{k\ell}(x), \mu_{k\ell}(x)$	0.23	0.24	0.24	0.22	0.19	0.23	0.23	0.22
$\delta_{k\ell}(x), \gamma_{k\ell}(x)$	-0.45	-0.46	-0.42	-0.42	-0.14	-0.23	-0.24	-0.26
$\lambda_{k\ell}(x), \gamma_{k\ell}(x)$	0.10	0.06	0.03	-0.01	0.04	0.10	0.11	0.07
$\psi_{k\ell}(x), \gamma_{k\ell}(x)$	0.31	0.31	0.27	0.24	0.25	0.29	0.31	0.29
$\delta_{k\ell}(x), \lambda_{k\ell}(x)$	-0.14	-0.15	-0.12	-0.11	-0.11	-0.14	-0.14	-0.08
$\psi_{k\ell}(x), \lambda_{k\ell}(x)$	0.85	0.83	0.81	0.76	0.75	0.80	0.75	0.70
$\psi_{k\ell}(x), \delta_{k\ell}(x)$	-0.19	-0.21	-0.17	-0.12	-0.11	-0.15	-0.15	-0.09
$\ln \gamma_{k\ell}(x), NPV_{k\ell}(x)$	0.75	0.75	0.75	0.74	0.48	0.58	0.55	0.59

- Strong pecuniary motive at short tenure. Less so at long tenure.
  - Moreso NPV of future earnings stream than job's current earnings.
- Negative correlation between preference and layoff into unemployment. ∃ reasonable stories of causality in both directions.
- Job finding exhibits both segmentation and preferences.

Preferences – in summary

- Workers are not indifferent about where they match.
- Preferences are stronger for higher worker wage-types. Similar to sorting mechanism in Bagger and Lentz (2019).
- $\gamma_{k\ell}(x)$ strongly positively correlated with NPV particularly for short tenure workers (0.7-0.8), but weaken with tenure. → non-wage factors matter.

Sorting

Measuring sorting

- Wage sorting: $\text{Cor}(a_k, b_\ell)$.

Measuring sorting

- Wage sorting: $\text{Cor}(a_k, b_\ell)$.
- Mutual information:

$$I(x) = \sum_{k,\ell} p(k,\ell|x) \ln\left(\frac{p(k,\ell|x)}{p(k|x)p(\ell|x)}\right)$$

  - Distance of observed matching distribution to independence.

Measuring sorting

- Wage sorting: $\text{Cor}(a_k, b_\ell)$.
- Mutual information:

$$I(x) = \sum_{k,\ell} p(k,\ell|x) \ln\left(\frac{p(k,\ell|x)}{p(k|x)p(\ell|x)}\right)$$

  - Distance of observed matching distribution to independence.
- Normalize by marginal entropy,

$$\tilde{I}(x) = \frac{I(x)}{\min\left[-\sum_k p(k|x)\ln\big(p(k|x)\big), -\sum_\ell p(\ell|x)\ln\big(p(\ell|x)\big)\right]}.$$

Wage sorting $\mathrm{Cor}(a_k(x), b_\ell(x))$

Type sorting $\bar{I}(k,\ell)$

Aggregate sorting – in summary

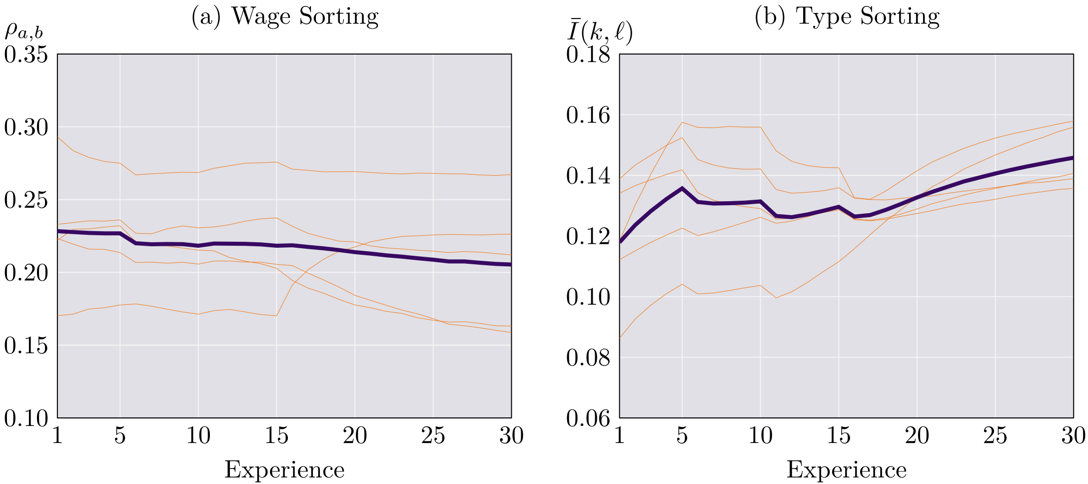
- Wage sorting:
  - Short tenure workers are more wage sorted than long tenure. Consistent w/ results on preferences
  - Wage sorting increasing over calendar time. Similar result in Bagger, Sørensen, and Vejlin (2013), Card, Heining, and Kline (2013).

- Wage sorting:
  - Short tenure workers are more wage sorted than long tenure. Consistent w/ results on preferences
  - Wage sorting increasing over calendar time. Similar result in Bagger, Sørensen, and Vejlin (2013), Card, Heining, and Kline (2013).
- Type sorting:
  - Long tenure more sorted than short tenure! Opposite of wage sorting. Again, consistent w/ preference study: Preferences intensify in tenure and age, but less correlated w/ wages.
  - No calendar time trend.

Sorting over the lifecycle

Synthetic cohorts

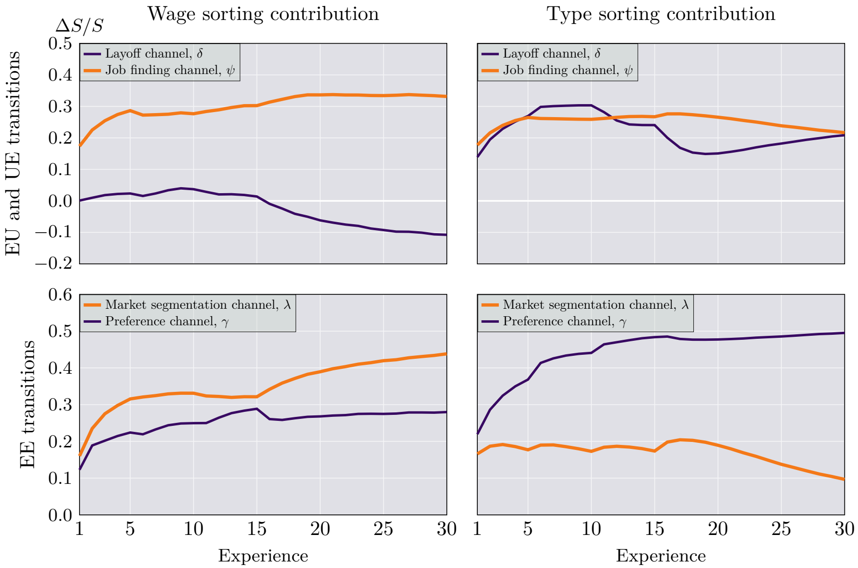
- To further investigate sorting determinants, study synthetic cohorts.
- Create cohorts starting with zero exp. and tenure.
- Simulate forward 30 years, holding calendar time fixed.


(a) Wage Sorting
$\rho_{a,b}$
0.35
0.30
0.25
0.20
0.15
0.10
1
5
10
15
20
25
30
Experience
(b) Type Sorting
$\bar{I}(k,\ell)$
0.18
0.16
0.14
0.12
0.10
0.08
0.06
1
5
10
15
20
25
30
Experience

Synthetic cohort counterfactuals

- Run counterfactuals on the mobility parameters by removing k, ℓ variations in:
  - Unemployment transitions
    - Layoff, δ
    - Reemployment, ψ
  - E-E transitions
    - Job preferences, γ
    - Segmentation, λ
- We also do counterfactuals where sorting is eliminated through removal of just one side (either k or ℓ variation). Results broadly similar.

Sorting contributions



Life cycle sorting – in summary

- Job preferences important sorting determinant - increasing w/ experience. Dominant on non-wage factors.
- Job finding channel key driver of both wage and type sorting at all experience levels.
- Market segmentation $\lambda_{k\ell'}$ drives a classical form of sorting via wage effects, also late in life. Unimportant for type sorting.

Sorting decomposition

Sorting decomposition

- To understand the interactions between the four channels, we turn each one off sequentially.
- 24 different orders of elimination.

SORTING DECOMPOSITION

- To understand the interactions between the four channels, we turn each one off sequentially.
- 24 different orders of elimination.
- Denote by $\bar{S} = \sum_{a,t} S_{at}$ aggregate sorting by age and time period.
- Let $\bar{S}_{-C}$ be the counterfactual aggregate sorting where channels in set C have been eliminated.

Sorting decomposition

- To understand the interactions between the four channels, we turn each one off sequentially.
- 24 different orders of elimination.
- Denote by $\bar{S} = \sum_{a,t} S_{at}$ aggregate sorting by age and time period.
- Let $\bar{S}_{-C}$ be the counterfactual aggregate sorting where channels in set C have been eliminated.
- Marginal impact of channel c given already eliminated channels C,

$$SC_{c,C} = \frac{\bar{S}_{-C} - \bar{S}_{-\{C,c\}}}{\bar{S}}.$$

Sorting decomposed

Elimination order	Wage sorting γ	δ	λ	ψ	Type sorting γ	δ	λ	ψ
1234	0.252	−0.068	0.450	0.366	0.446	−0.084	0.342	0.296
1243	0.252	−0.068	0.119	0.697	0.446	−0.084	0.180	0.458
1324	0.252	−0.173	0.555	0.366	0.446	−0.037	0.294	0.296
1342	0.252	0.207	0.555	−0.014	0.446	0.144	0.294	0.115
1423	0.252	0.461	0.119	0.169	0.446	0.156	0.180	0.218
1432	0.252	0.207	0.372	0.169	0.446	0.144	0.192	0.218
2134	0.209	−0.024	0.450	0.366	0.145	0.217	0.342	0.296
2143	0.209	−0.024	0.119	0.697	0.145	0.217	0.180	0.458
2314	0.457	−0.024	0.202	0.366	0.281	0.217	0.206	0.296
2341	0.179	−0.024	0.202	0.644	0.089	0.217	0.206	0.488
2413	0.031	−0.024	0.119	0.874	0.005	0.217	0.180	0.598
2431	0.179	−0.024	−0.028	0.874	0.089	0.217	0.096	0.598
3124	0.456	−0.173	0.351	0.366	0.573	−0.037	0.168	0.296
3142	0.456	0.207	0.351	−0.014	0.573	0.144	0.168	0.115
3214	0.457	−0.173	0.351	0.366	0.281	0.254	0.168	0.296
3241	0.179	−0.173	0.351	0.644	0.089	0.254	0.168	0.488
3412	0.381	0.207	0.351	0.062	0.580	0.144	0.168	0.108
3421	0.179	0.409	0.351	0.062	0.089	0.634	0.168	0.108
4123	0.120	0.461	0.119	0.301	0.414	0.156	0.180	0.250
4132	0.120	0.207	0.372	0.301	0.414	0.144	0.192	0.250
4213	0.031	0.549	0.119	0.301	0.005	0.565	0.180	0.250
4231	0.179	0.549	−0.028	0.301	0.089	0.565	0.096	0.250
4312	0.381	0.207	0.111	0.301	0.580	0.144	0.026	0.250
4321	0.179	0.409	0.111	0.301	0.089	0.634	0.026	0.250
Average	0.245	0.129	0.256	0.369	0.300	0.214	0.183	0.302
Std dev	0.124	0.237	0.165	0.247	0.199	0.198	0.077	0.140

Sorting decomposition – in summary

- Job preferences are an important contributor to both wage and type sorting ($25 - 30\%$)).
- Together, job preferences (γ) and segmentation (λ) contribute some 70%.
- Generally, channels complement each other in sorting.
  - Preferences and layoffs are a particularly strong pairwise interaction.
  - Layoff channel comes to prominence when job finding channel already eliminated.
- Layoff channel is relatively insignificant to wage sorting, but matters to type sorting. Although, see its interaction w/ job finding.

Conclusion

Big picture take away

- Workers perceive significant value dispersion across employer types.
  - Preference strength increasing in worker wage type.
  - Preferences related substantially to wages (in particular to NPV), but non-wage factors important.
  - Preferences contribute substantially to sorting.

Big picture take away

- Workers perceive significant value dispersion across employer types.
  - Preference strength increasing in worker wage type.
  - Preferences related substantially to wages (in particular to NPV), but non-wage factors important.
  - Preferences contribute substantially to sorting.
- Sorting is increasing over life cycle. Early on wages, later less so and in particular driven by job preference channel.

Big picture take away

- Workers perceive significant value dispersion across employer types.
  - Preference strength increasing in worker wage type.
  - Preferences related substantially to wages (in particular to NPV), but non-wage factors important.
  - Preferences contribute substantially to sorting.
- Sorting is increasing over life cycle. Early on wages, later less so and in particular driven by job preference channel.
- Methodology: Preferences revealed (along w/ rest of mobility model) in setup where latent types are identified through both wages and mobility.

Appendix

Recursive Job Value

- With some offense to the AR(1) assumption of wages, let $y_{k\ell}(x)$ be the utility flow of the match.
- Assuming each event arrival is exclusive, match value can be stated recursively by,

$$V_{k\ell}(x) = y_{k\ell}(x) + \beta\Big[\nu_{kx}\sum_{\ell'=1}^{L}\lambda_{k\ell'}(x')\ln\big(\exp(V_{k\ell}(x')/\nu_{kx}) + \exp(V_{k\ell'}(x')/\nu_{kx})\big) + \nu_{kx}\lambda_k(x)\vartheta$$
$$+\,\delta_{k\ell}(x)V_{k0}(x') + \big(1-\lambda_k(x)-\delta_{k\ell}(x)\big)V_{k\ell}(x')\Big],$$

where

- $\lambda_k(x) = \sum_{\ell'=1}^{L}\lambda_{k\ell'}(x)$
- $\vartheta \approx .577$ is the Euler-Mascheroni constant.
- ν_{kx} is the variance parameter of the random utility or mobility cost realization
- β is the discount factor.

«

Likelihood given firm classification

- The complete likelihood of worker i's history and that s/he is type k is

$$\mathcal{L}_i(k;\beta,F) = \frac{m^0_{k\ell_{i1}}(x)\pi(z^w|k)\pi(z^f|\ell_{i1})}{\#(\ell_{i1}|F)} \prod_{t=1}^{T-1} M_{k\ell_{it}\neg}(x_{it})^{1-D_{it}} \left(\frac{M_{k\ell_{it}\ell_{i,t+1}}(x_{it})}{\#(\ell_{i,t+1}|F)}\right)^{D_{it}} \times$$
$$\prod_{t=1}^{T} f^0_{k\ell_{it}}(w_{it}|x_{it})^{D_{i,t-1}} f_{k\ell_{it}}(w_{it}|x_{it})^{1-D_{i,t-1}}.$$

- We use a modified EM algorithm to obtain estimate 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:

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

- Transition probabilities:

$$(\lambda, \gamma, \phi, \delta)^{(m+1)} = \arg\max_{M} \sum_{i,k} p_i(k; \beta^{(m)}, F)$$
$$\times \left(\sum_{t} \left[(1 - D_{it}) \ln M_{k\ell_{it}\neg}(x_{it}) + D_{it} \ln M_{k\ell_{it}\ell_{i,t+1}}(x_{it}) \right] \right)$$

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

«

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