A Hidden Markov Model of Wages and Employment Mobility with Worker and Firm Heterogeneity

Evidence from Italian (and Danish) Data

Long Hong   Rasmus Lentz   Jean-Marc Robin

NBER SI 2025 – The Micro and Macro Perspectives of the Aggregate Labor Market
July 14-18, 2025

Introduction

Wage and job mobility determinants

- How important are type dynamics to a worker's wage and job mobility outcomes?

Wage and job mobility determinants

- How important are type dynamics to a worker's wage and job mobility outcomes?
- We develop a rich framework for the flexible identification of determinants of wage and job mobility outcomes that includes the identification of,
  - Latent worker type heterogeneity
  - Latent firm type heterogeneity
  - A Markov process for worker type dynamics that also depends on the worker's current employer's type.

Wage and job mobility determinants

- How important are type dynamics to a worker's wage and job mobility outcomes?
- We develop a rich framework for the flexible identification of determinants of wage and job mobility outcomes that includes the identification of,
  - Latent worker type heterogeneity
  - Latent firm type heterogeneity
  - A Markov process for worker type dynamics that also depends on the worker's current employer's type.
- We make significant progress in the classification of firms by use of variational EM methods.

Preliminary results on Italian (Veneto) data

- Identify a combination of latent permanent worker type heterogeneity with type dynamics within each permanent type.

Preliminary results on Italian (Veneto) data

- Identify a combination of latent permanent worker type heterogeneity with type dynamics within each permanent type.
- Worker wage effect variance dominant explanation of wage variance. Firm wage effects much smaller contribution and less than the contribution from sorting.

Preliminary results on Italian (Veneto) data

- Identify a combination of latent permanent worker type heterogeneity with type dynamics within each permanent type.
- Worker wage effect variance dominant explanation of wage variance. Firm wage effects much smaller contribution and less than the contribution from sorting.
- Human capital and Search capital growth:
  - Wage growth by experience primarily explained with worker wage effect growth. Firm wage effect growth plays a non-negligible role, especially for women.

Preliminary results on Italian (Veneto) data

- Identify a combination of latent permanent worker type heterogeneity with type dynamics within each permanent type.
- Worker wage effect variance dominant explanation of wage variance. Firm wage effects much smaller contribution and less than the contribution from sorting.
- Human capital and Search capital growth:
  - Wage growth by experience primarily explained with worker wage effect growth. Firm wage effect growth plays a non-negligible role, especially for women.
- Worker wage effect (human capital) growth varies substantially across firm types:
  - Higher wage firms grow a worker's wage effect by more.
  - Supermodularity: Higher wage firms grow higher wage workers' wage effects by more.

Model

Matches and agent heterogeneity

- Adds HMM to Bonhomme, Lamadon, and Manresa (2019) and Lentz, Piyapromdee, and Robin (2023).
- A job is a match between a worker and a firm.
- A worker is at any point in time characterized by latent type $k \in \{1, \dots, K\}$. No ordering imposed.
- A firm is characterized by (ℓ, θ) where
  - latent type $\ell \in \{1, \dots, L\}$. No ordering imposed.
  - θ is probability that a worker meets the firm conditional on meeting a type ℓ firm.
- 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.
- Non-employment treated as match with firm $j = 0$ with $(\ell, \theta)_{j=0} = (0, 1)$.

Worker type transitions

- A worker's type k follows a hidden Markov process.
- Each spell-year a type k worker matched with type ℓ draws a type realization from $A(k' \mid k, \ell)$.
- A spell-year ends when the calendar year or the match ends, whichever comes first.

Job Mobility

- Each period, a type k worker currently with a type ℓ firm moves to a type ℓ' firm with probability $M_{k\ell\ell'}$.
- By implication, probability of staying is $M_{k\ell\neg} = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}$.

Initialization

- Initial worker type distribution, π^w
- Initial firm type distribution, π^f.
- Initial match distribution, $m(k, \ell)$, where $\sum_{\ell=0}^{L} m(k, \ell) = 1$.

- Match wages are log-normally distributed.
- Specifically, log wage, w, is distributed according to,

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

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

Data and Estimation

Data

- Italian register data, 1982-2001.
- Data on monthly wages, worker and employer IDs.
- Observable worker characteristics: Age, sex, coarse occupation description.
- As in Lentz, Piyapromdee, and Robin (2023), more observable characteristics can be included in analysis. Danish data are richer in this respect.
- For the Italian data, a period is a month. Wages are aggregated to the spell-year level.

Data summary

	(1) Mean	(2) S.D.	(3) Median
Daily wage	137.67	281.24	
Age	33.00	8.16	32
Wage chg cond on move	0.10	2.67	0
Movers	0.18	0.38	
Female	0.34	0.47	
Obs in Veneto	0.71	0.45	
Firm-year level stats for firms In Veneto			
Firm size	10.82	65.84	3.00
Movers per firm-year	1.57	13.62	0.00
Frac of movers per firm-year	0.17	0.29	0.00
Person-year observations		23,733,747	
Number of workers		2,433,225	
Number of firms		630,698	

Variance decompositions

$$\ln w_{it} = \alpha_i + \psi_{j(i,t)} + \beta X_{it} + u_{it}$$

	AKM Plug-in	KSS Leave pers-yr	KSS Leave match
$Var(\alpha)$	47.0%	40.1%	38.1%
$Var(\psi)$	20.1%	17.9%	16.9%
$2 \times Cov(\alpha, \psi)$	4.7%	8.0%	
$Total(\alpha, \psi)$	71.8%	66.0%	64.3%
$Corr(\alpha, \psi)$	0.08	0.15	0.18
Variance of y		0.139	

Bipartite Degree Corrected Stochastic Block Model.

- Worker i's history comprises S_i spell-year observations,

$$
\begin{aligned}
X_i(1) &= \big(Y_i(1), W_i(1), D_i(s), Y_i(2)\big) \\
X_i(s) &= \big(W_i(s), D_i(s), Y_i(s+1)\big),\ s = 2, \ldots, S_i - 1 \\
X_i(S_i) &= \big(W_i(S_i), D_i(S_i)\big),
\end{aligned}
$$

  where,
  - Indicator $Y_{ij}(s) = 1$ if worker i matched with firm j in spell-year s. $\sum_j Y_{ij}(s) = 1$.
  - $D_i(s)$ is duration of spell-year s.
  - $W_i(s)$ is wage in spell-year s.

Bipartite Degree Corrected Stochastic Block Model.

- Worker i's history comprises S_i spell-year observations,

$$
\begin{aligned}
X_i(1) &= \big(Y_i(1), W_i(1), D_i(s), Y_i(2)\big) \\
X_i(s) &= \big(W_i(s), D_i(s), Y_i(s+1)\big),\ s = 2, \ldots, S_i - 1 \\
X_i(S_i) &= \big(W_i(S_i), D_i(S_i)\big),
\end{aligned}
$$

  where,
  - Indicator $Y_{ij}(s) = 1$ if worker i matched with firm j in spell-year s. $\sum_j Y_{ij}(s) = 1$.
  - $D_i(s)$ is duration of spell-year s.
  - $W_i(s)$ is wage in spell-year s.
- Worker latent types (communities):
  - $Z^{w}_{ik}(s) = 1$ if worker i is type k in spell s. $\sum_k Z^{w}_{ik}(s) = 1$.
- Firm latent types (communities):
  - $Z^{f}_{j\ell} = 1$ if firm j is type ℓ. $\sum_\ell Z^{f}_{j\ell} = 1$.
  - Degree $\theta_{j\ell}$.

Variational Expectation Maximization (VEM)

- Goal: Maximize incomplete likelihood $\mathcal{L}(X; b)$. Integrates over latent types $Z = (Z^w, Z^f)$.
- EM algorithm does this through iterative maximization of the expected *complete log likelihood*, $b^{m+1} = \arg\max_b [\sum_Z R^m(Z) \ln \mathcal{L}(X, Z; b)]$, where $R^m(Z) = \mathcal{L}(Z \mid X; b^m)$ is Z posterior given data and model parameters b^m. Application of Minorization-Maximization algorithm.

Variational Expectation Maximization (VEM)

- Goal: Maximize incomplete likelihood $\mathcal{L}(X; b)$. Integrates over latent types $Z = (Z^w, Z^f)$.
- EM algorithm does this through iterative maximization of the expected *complete log likelihood*, $b^{m+1} = \arg\max_b [\sum_Z R^m(Z) \ln \mathcal{L}(X, Z; b)]$, where $R^m(Z) = \mathcal{L}(Z \mid X; b^m)$ is Z posterior given data and model parameters b^m. Application of Minorization-Maximization algorithm.
- In our case, not feasible to obtain $\mathcal{L}(Z \mid X; b^m)$. Firm type posterior dependence.

Variational Expectation Maximization (VEM)

- Goal: Maximize incomplete likelihood $\mathcal{L}(X; b)$. Integrates over latent types $Z = (Z^w, Z^f)$.
- EM algorithm does this through iterative maximization of the expected *complete log likelihood*, $b^{m+1} = \arg\max_b[\sum_Z R^m(Z) \ln \mathcal{L}(X, Z; b)]$, where $R^m(Z) = \mathcal{L}(Z \mid X; b^m)$ is Z posterior given data and model parameters b^m. Application of Minorization-Maximization algorithm.
- In our case, not feasible to obtain $\mathcal{L}(Z \mid X; b^m)$. Firm type posterior dependence.
- *VEM algorithm*:
  - Pseudo E step: Given feasible set $\mathcal{R}$, choose $\hat{R}^m$ to minimize distance to $\mathcal{L}(Z \mid X; b^m)$.
  - M step: $b^{m+1} = \arg\max_b[\sum_Z \hat{R}^m(Z) \ln \mathcal{L}(X, Z; b)]$
  - Update b^m with b^{m+1}. Repeat until convergence.

VEM FOR OUR MODEL

- Choice of feasible set $\mathcal{R}$,
  - Force independence between worker and firm types, $R(Z) = R^w(Z^w)R^f(Z^f)$.
  - Force independence between firms priors, τ_j, where $\sum_\ell \tau_{j\ell} = 1$,

$$R^f(Z^f) = \prod_{j=1}^{J} R_j^f(Z_j^f) = \prod_{j=1}^{J} \tau_{j\ell}^{Z_{j\ell}^f}.$$

VEM FOR OUR MODEL

- Choice of feasible set $\mathcal{R}$,
  - Force independence between worker and firm types, $R(Z) = R^w(Z^w)R^f(Z^f)$.
  - Force independence between firms priors, τ_j, where $\sum_\ell \tau_{j\ell} = 1$,

$$R^f(Z^f) = \prod_{j=1}^{J} R^f_j(Z^f_j) = \prod_{j=1}^{J} \tau_{j\ell}^{Z^f_{j\ell}}.$$

- With that, the pseudo E-step is tractable (and quite fast), and M-step remains a simple set of analytical solutions for model parameters based on first order conditions.
  - Baum-Welch algorithm remains available for the determination of worker type marginals, $\zeta_{ik}(s) = \Pr\big(Z^w_{ik}(s) = 1\big)$ and $\zeta_{ikk'}(s) = \Pr\big(Z^w_{ik}(s-1) = 1 \text{ and } Z^w_{ik'}(s) = 1\big)$.
  - τ_j follows from sparse system of first order conditions (minimize distance to $\mathcal{L}(Z \mid X; b^m)$).

VEM FOR OUR MODEL

- Choice of feasible set $\mathcal{R}$,
  - Force independence between worker and firm types, $R(Z) = R^w(Z^w)R^f(Z^f)$.
  - Force independence between firms priors, τ_j, where $\sum_\ell \tau_{j\ell} = 1$,

$$R^f(Z^f) = \prod_{j=1}^{J} R_j^f(Z_j^f) = \prod_{j=1}^{J} \tau_{j\ell}^{Z_{j\ell}^f}.$$

- With that, the pseudo E-step is tractable (and quite fast), and M-step remains a simple set of analytical solutions for model parameters based on first order conditions.
  - Baum-Welch algorithm remains available for the determination of worker type marginals, $\zeta_{ik}(s) = \Pr\big(Z_{ik}^w(s) = 1\big)$ and $\zeta_{ikk'}(s) = \Pr\big(Z_{ik}^w(s-1) = 1 \text{ and } Z_{ik'}^w(s) = 1\big)$.
  - τ_j follows from sparse system of first order conditions (minimize distance to $\mathcal{L}(Z \mid X; b^m)$).
- Concentration: When firm priors τ_j are fully concentrated (full mass on single type), the assumptions of posterior independence no longer restrictive, and $\mathcal{R}$ includes $\mathcal{L}(Z \mid X; b^m)$.
  - Links back to Lentz, Piyapromdee, and Robin (2023) CEM estimation where we search over hard firm classifications.

Estimator performance

- Identification proof strategy similar to BLM and LPR. Sufficient to have 3 periods.
- We have demonstrated that estimator can reliably capture true model parameters on simulated data. More systematic work still to be done.
- For a sense of speed, in the following a single estimate takes 5-10 minutes for a single 128 cores machine. We are showing the best of 500 restarts.

Preliminary Results

Types

- $K = 3 \times 3 = 9$ and $L = 5$.
- Observed characteristics: $z \in 1, \ldots, 8$. Entry age by sex. Age: (21-27), (27-33), (34-40), (41-50).
  - Enter through initial worker type realization distribution, $\pi_w(z)$ and $m(k, \ell | z)$.
- 3 permanent types (blocks). Each block has type dynamics characterized by A_ℓ^b with 3 states.
  - Impose block diagonal structure on type transition matrix,

$$A(k, k', \ell) = \begin{pmatrix} A_\ell^1 & 0 & 0 \\ 0 & A_\ell^2 & 0 \\ 0 & 0 & A_\ell^3 \end{pmatrix},$$

where A_ℓ^b is a (3×3) matrix, $b \in \{1, 2, 3\}$.

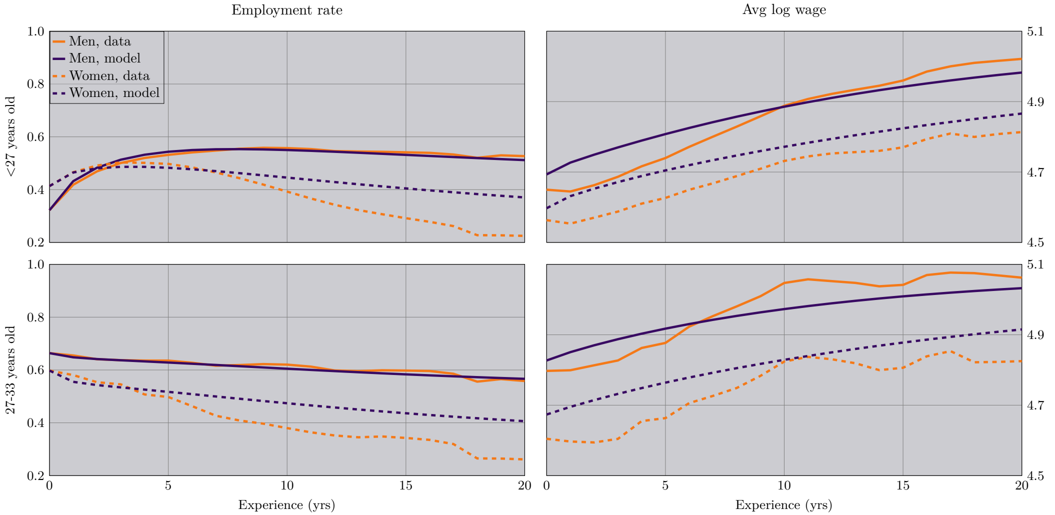
- For a given worker, model implies an overall Markov process in (k, ℓ). 54 states.
- Given spell-year time structure and monthly frequency, step state forward 11 times according to,

$$\Pr(k', \ell' | k, \ell) = \begin{cases} M_{k\ell\ell'} A(k'|k, \ell) & \text{if } k' \neq k \\ M_{k\ell\neg} + M_{k\ell\ell} A(k|k, \ell) & \text{if } k' = k \text{ and } \ell' = \ell \\ M_{k\ell\ell'} A(k|k, \ell) & \text{otherwise.} \end{cases}$$

- For end-of-year, step forward according to,

$$\Pr(k', \ell' | k, \ell) = \begin{cases} [M_{k\ell\neg} + M_{k\ell\ell}] A(k'|k, \ell) & \text{if } \ell' = \ell \\ M_{k\ell\ell'} A(k'|k, \ell) & \text{otherwise.} \end{cases}$$

Modelfit to employment and wage dynamics



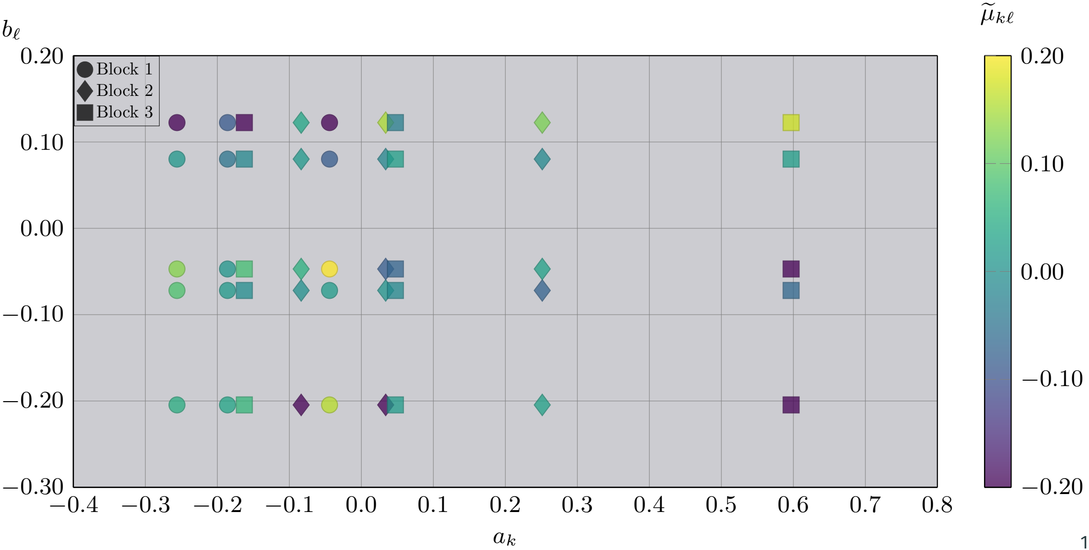
Wage labels

- Wage types assigned through the linear projection:

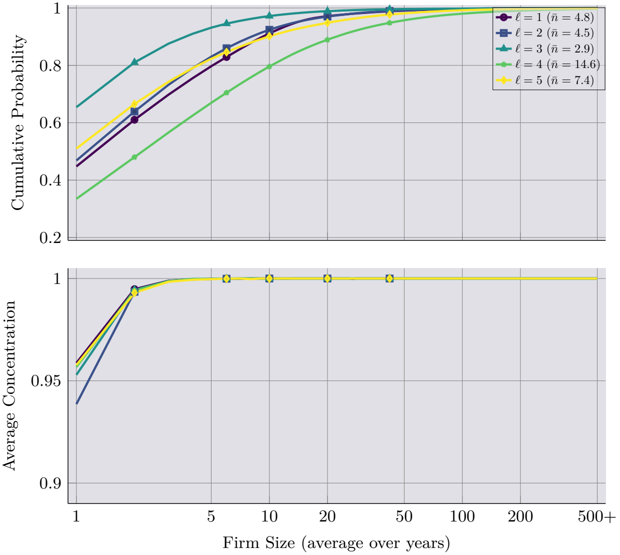
$$\mu_{k\ell} = \bar{\mu} + a_k + b_\ell + \tilde{\mu}_{k\ell}.$$

- a_k worker wage type.
- b_ℓ firm wage type.
- Order worker types by average block a_k, then by a_k. Low to high.
- Order firm types by b_ℓ. Low to high.

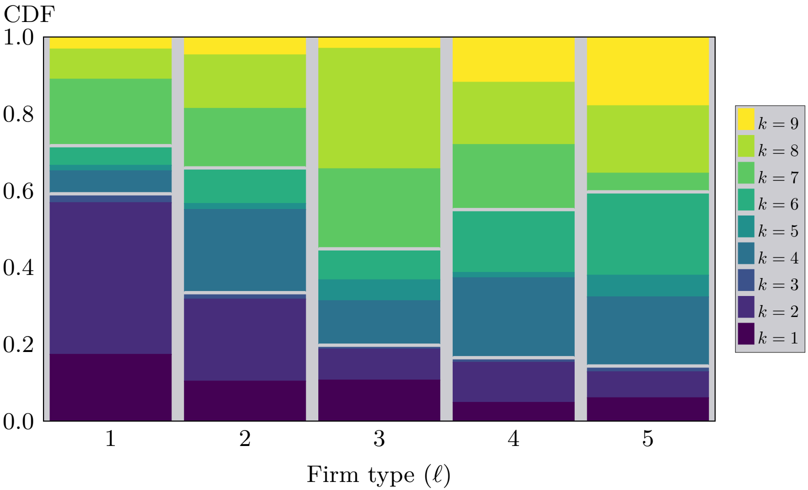
Wage labels and residuals



Firm types. Concentration (max τ), and size distribution by type



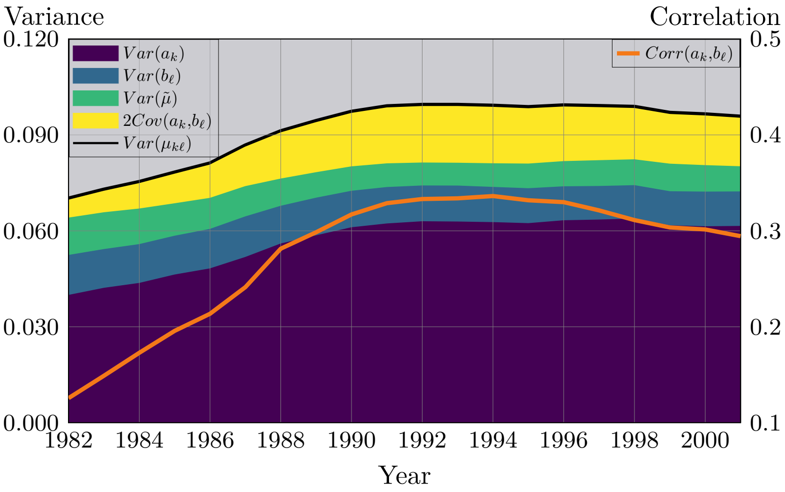
Firm type conditional worker type distribution



Wage variance decomposition

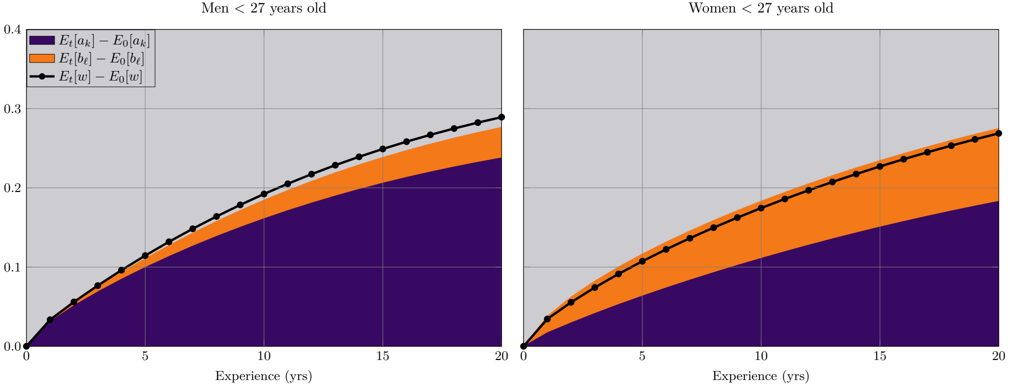
$Var(\mu_{k\ell})$		0.092
Decomposed into:		
$\quad Var(a_k)$	0.625	
$\quad Var(b_\ell)$	0.123	
$\quad Var(\tilde{\mu}_{k\ell})$	0.094	
$\quad 2Cov(a_k, b_\ell)$	0.158	
$Corr(a_k, b_\ell)$		0.285

Wage variance decomposition over time



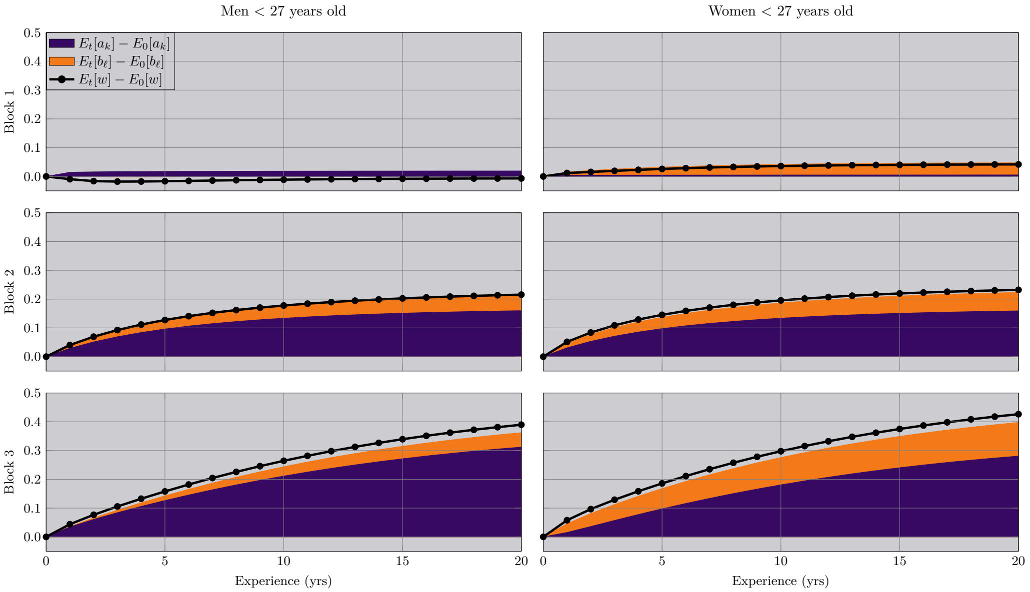
Cohort wage growth

- Decompose a cohort's wage growth by experience into worker and firm wage effect growth (leaving out non-linearity change). Loosely, think human capital vs search capital growth.

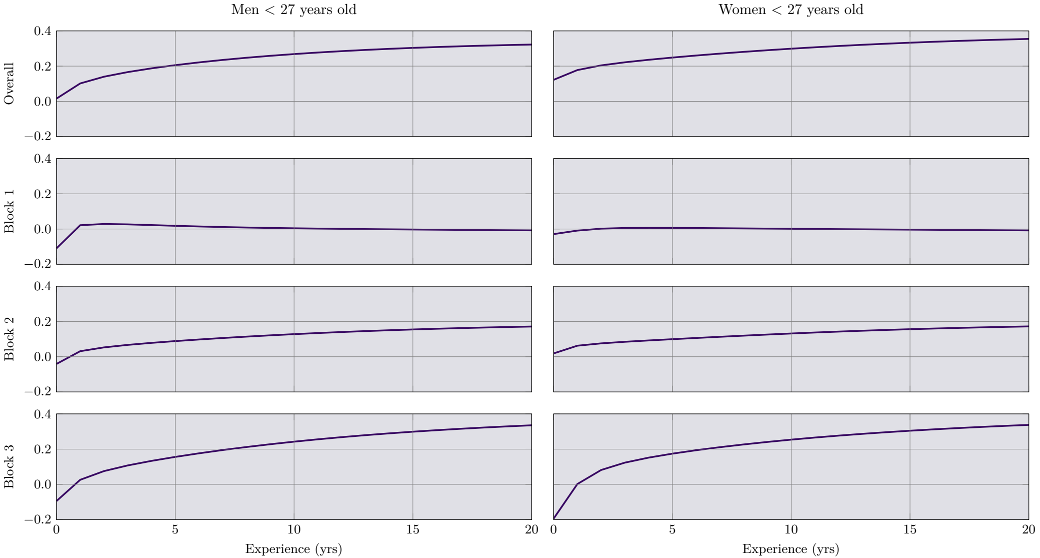


- Especially, for men, wage growth primarily explained through own wage effect growth.

Cohort wage growth, by permanent types (blocks)



Cohort wage sorting, overall and by permanent types

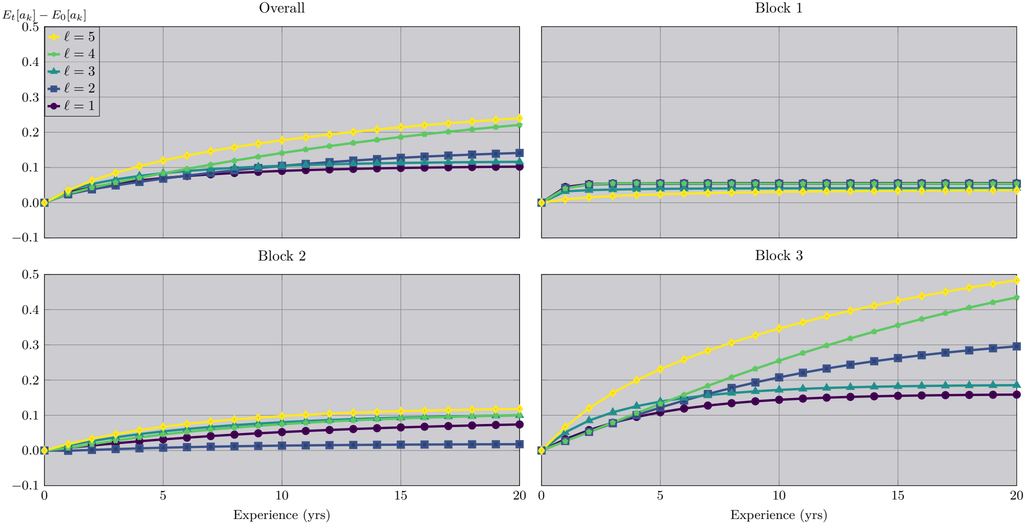


a_k GROWTH HETEROGENEITY ACROSS FIRM TYPES

- There is substantial variation across firm types in how much they grow a worker's wage effect. Think training heterogeneity by firm type.
- Perform counterfactual of continued employment with a given firm type.

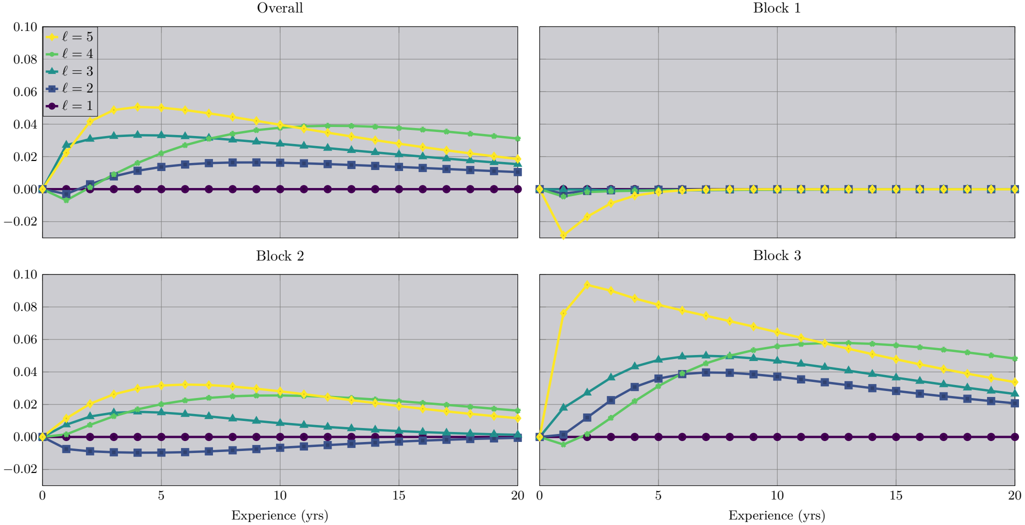
Counterfactual: Employment with a fixed firm type. Men < 27 years old

a_k growth, $E_t[a_k] - E_0[a_k]$



Counterfactual: Impact of first job on a_k path. Men < 27 years old

$$E_t[a_k \mid \text{first job with } \ell] - E_t[a_k \mid \text{first job with } \ell = 1]$$



Concluding thoughts

- Worker type variation dominant contribution to static wage variance.
- Growth in worker wage type dominant source of overall wage growth.
- Firm heterogeneity seemingly important determinant in worker type dynamics variability.
  - Higher type firms grow worker effects by more.
  - Furthermore, higher type firms grow higher type workers by more.
- We demonstrate VEM as an attractive method for worker and firm classification.

Appendix

Complete information log likelihood

- The complete likelihood where $Z = (Z^w, Z^f)$ is known,

$$\ln \mathcal{L}\left(X, Z^f, Z^w\right) = \ln \mathcal{L}\left(Z^f\right) + \sum_{i=1}^{I} \ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right),$$

  where,

$$\ln \mathcal{L}\left(Z^f\right) = \sum_{j=1}^{J}\sum_{\ell=1}^{L} Z_{j\ell}^f \ln \pi_\ell^f .$$

Worker i firm classification conditional complete log likelihood

- Worker i's complete log likelihood is,

$$\ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right) = \sum_{j=0}^{J} Y_{ij}(1) \sum_{\ell=1}^{L} Z_{j\ell}^f \sum_{k=1}^{K} Z_{ik}^w(1) \ln\left[\pi_k^w m_{k\ell} \theta_{j\ell}\right]$$
$$+ \sum_{s=2}^{S_i} \sum_{k=1}^{K} Z_{ik}^w(s-1) \sum_{k'=1}^{K} Z_{ik'}^w(s) \ln \alpha_{kk'}(s|Z^f) + \sum_{s=2}^{S_i} \sum_{k=1}^{K} Z_{ik}^w(s) \ln \beta_k(s|Z^f),$$

where for $s = 2, \ldots, S_i$,

$$\ln \alpha_{kk'}(s|Z^f) = \sum_{j=1}^{J} Y_{ij}(s-1) \sum_{\ell=0}^{L} Z_{j\ell}^f \ln A_{k\ell k'}$$

$$\ln \beta_k(s|Z^f) = \sum_{j=0}^{J} Y_{ij}(s) \sum_{\ell=1}^{L} Z_{j\ell}^f \Bigg(\mathbf{1}\{j \neq 0\} \ln f_{k\ell}(W_i(s)) + D_i(s) \ln M_{k\ell\neg}$$
$$+ \sum_{j'=0}^{J} Y_{ij'}(s+1) \sum_{\ell'=0}^{L} Z_{j'\ell'}^f \mathbf{1}\{j' \neq j\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right] \Bigg).$$

Variational EM

- For model parameters b and a probability distribution $R(Z)$, define a lower bound on the incomplete log likelihood, $\mathcal{J}(R, X; b)$ using the Kullback-Leibler divergence,

$$\begin{aligned}
\mathcal{J}(R, X; b) &= \ln \mathcal{L}(X; b) - D_{KL}\left(R \,\|\, \mathcal{L}(Z \mid X; b)\right) \\
&= \ln \mathcal{L}(X; b) - \sum_Z R(Z) \ln \left(\frac{R(Z)}{\mathcal{L}(Z \mid X; b)} \right) \\
&= \sum_Z R(Z) \ln \mathcal{L}(X, Z; b) + \mathcal{H}(R),
\end{aligned}$$

  where $\mathcal{H}(R) = -\sum_Z R(Z) \ln R(Z)$ is the R distribution entropy.
- If $\mathcal{L}(Z \mid X; b)$ is tractable, then EM algorithm is available to maximize incomplete likelihood.
  - Uses $R^*(Z; b) = \mathcal{L}(Z \mid X; b)$ in which case $\mathcal{J}(R^*(Z; b^0), X; b)$ becomes a minorization of $\ln \mathcal{L}(X; b)$ in b^0.
- VEM: Given feasible set $\mathcal{R}$, choose R to maximize $\mathcal{J}$,

$$\widehat{R} = \arg\max_{R \in \mathcal{R}} \mathcal{J}(R, X) = \arg\max_{R \in \mathcal{R}} \sum_Z R(Z) \ln \mathcal{L}(Z \mid X) - \sum_Z R(Z) \ln R(Z)$$

Posterior dependence example

- Firms A and B connected by worker i through move from A to B. Ignore worker classification.
- 2 firm types. Worker i mobility matrix,

$$M = \begin{pmatrix} M_{11} & 0 \\ 0 & M_{22} \end{pmatrix}.$$

- The data conditional classification prob has $\mathcal{L}(Z^f_{A1} = 1, Z^f_{B2} = 1 \mid X) = \mathcal{L}(Z^f_{A2} = 1, Z^f_{B1} = 1 \mid X) = 0$.
- Our VEM imposes,

$$R^f(Z^f_{A1} = 1, Z^f_{B2} = 1) = \tau_{A1}\tau_{B2}$$
$$R^f(Z^f_{A1} = 1, Z^f_{B2} = 1) = \tau_{A2}\tau_{B1}$$