A Distributional Framework for Matched Employer Employee Data

Stéphane Bonhomme Thibaut Lamadon Elena Manresa
University of Chicago MIT Sloan

PRELIMINARY AND INCOMPLETE

NBER Summer Institute 2015
Labor Studies

Worker and firm heterogeneity

• Important questions: firm/worker sorting and optimal allocations, sources of wage inequality...

• For more than 20 years, researchers have relied on **matched data** that follow workers across firms over time.

• Matched panel datasets are useful to allow for both unobserved worker and firm heterogeneity.

Fixed-effects regression

• Abowd, Kramarz and Margolis (1999) estimated:

$$log\, wage = worker\, FE + firm\, FE + covariates + error\, term$$

on French matched employer employee data.

• This allows to document the association $cov(worker\, FE, firm\, FE)$, and more generally the contributions of workers and firms to wage dispersion.

• Widely applied method, in labor (for example Card, Heining and Kline 2013) and outside (schools, hospitals, cities...).

Fixed-effects regression (cont.)

1. Additive model (in logs). Hence a very specific form of **complementarity** between worker and firm characteristics.

-Theoretical and structural literature on sorting models (Becker 1974, Shimer and Smith 2000, Eeckhout and Kircher 2011, among others).

2. In panels with **limited job mobility**, estimates based on a fixed-effects approach may be biased (Andrews *et al.* 2008).

This paper

- Estimate **wage distributions** conditional on worker and firm heterogeneity, and estimate **worker composition** in different firms.

- The framework is consistent with a number of **theoretical** sorting models.

- **Discrete** heterogeneity: “grouped fixed-effects” for firms, and “correlated random-effects” for workers.

- Show nonparametric identification. 3-step estimation method. Apply the method to Swedish employer-employee data.

- Key source of variation: wages of **job movers** (as in fixed-effects regressions).

Plan of the talk

- Framework
- Identification and Estimation
- Data and empirical results (preliminary)

Framework

Heterogeneity and wages in period 1

• Workers indexed by i with discrete **types** $\omega(i) \in \{1, ..., K\}$, and firms indexed by j with discrete **classes** $f(j) \in \{1, ..., L\}$.

• Let j_{it} denote the identifier of the firm where i works at time t.

• The proportion of type-k workers (that is, $\omega(i) = k$) in a class-ℓ firm (that is, $f(j_{i1}) = \ell$) is $\boldsymbol{\pi_k(\ell)}$.

• The cdf of log wages Y_{i1} for a type-k worker in a class-ℓ firm is $\boldsymbol{F_{k\ell}(y)}$.

• Here we abstract from worker- and firm-level covariates.

Job mobility (2-periods model)

• <u>Period 1:</u>

-A type-k worker in a class-ℓ firm draws a wage Y_{i1} from $F_{k\ell}(y)$.

• <u>Period 2:</u>

-The worker moves to a class-ℓ' firm with a probability that depends on k and ℓ, **not on** $\boldsymbol{Y_{i1}}$.

-If he moves, the worker draws a wage Y_{i2} from a distribution $G_{k\ell'}(y')$ that depends on k and ℓ', **not on** $\boldsymbol{(\ell, Y_{i1})}$.

• Two assumptions: 1) mobility is driven by types/classes, 2) serial independence upon job change.

Job mobility (4-periods model)

• Periods 1 and 2: A type-k worker in a class-ℓ firm draws wages (Y_{i1}, Y_{i2}) from a bivariate distribution that depends on (k, ℓ).

• Period 3:

-The worker moves to a class-ℓ' firm with a probability that depends on k, ℓ and Y_{i2}, **not on $\boldsymbol{Y_{i1}}$**.

-If he moves, the worker draws a wage Y_{i3} from a distribution that depends on k, ℓ', ℓ, Y_{i2}, **not on $\boldsymbol{Y_{i1}}$**.

• Period 4: The worker then draws a wage Y_{i4} from a distribution that depends on k, ℓ', Y_{i3}, **not on $\boldsymbol{(\ell, Y_{i2}, Y_{i1})}$**.

Statistical implications

- <u>2-periods model:</u> if $m_i = \mathbf{1}\{j_{i1} \neq j_{i2}\}$ denotes job mobility, then

$$Y_{i1} \text{ and } Y_{i2} \text{ are conditionally independent given } (m_i = 1, k, \ell, \ell').$$

Type proportions of job movers are denoted as $p_k(\ell, \ell')$.

- <u>4-periods model:</u> if $m_i = \mathbf{1}\{j_{i2} \neq j_{i3}\}$ denotes job mobility between periods 2 and 3, then

$$Y_{i1} \text{ and } Y_{i4} \text{ are conditionally independent given } (m_i = 1, k, \ell, \ell', Y_{i2}, Y_{i3}).$$

Main restrictions (2-periods model)

- In the 2-periods model we have:

$$\Pr\left[Y_{i1} \leq y\,,\, Y_{i2} \leq y' \,|\, m_i = 1, f(j_{i1}) = \ell, f(j_{i2}) = \ell'\right] = \sum_{k=1}^{K} p_k\left(\ell, \ell'\right) F_{k\ell}(y) G_{k\ell'}(y'), \qquad \text{(MOV)}$$

$$\Pr\left[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell\right] = \sum_{k=1}^{K} \pi_k\left(\ell\right) F_{k\ell}(y). \qquad \text{(CROSS)}$$

- Firm classes $f(j)$ can be recovered using (MOV) and/or (CROSS).

- We show how to recover $F_{k\ell}$, $G_{k\ell'}$, and $p_k(\ell, \ell')$ from (MOV) using job movers, and $\pi_k(\ell)$ from (CROSS) using the first cross-section.

Main restrictions (4-periods model)

- In the more general 4-periods model we have:

$$\Pr\left[Y_{i1} \leq y_1, Y_{i4} \leq y_4 \,|\, m_i = 1, Y_{i2} = y_2, Y_{i3} = y_3, f(j_{i2}) = \ell, f(j_{i3}) = \ell'\right]$$
$$= \sum_{k=1}^{K} p_k\left(y_2, y_3, \ell, \ell'\right) F_{k\ell}(y_1|y_2) G_{k\ell'}(y_4|y_3). \tag{MOV-ENDO}$$

- **Similar structure** as in the setup with exogenous mobility, one period further away on each side.

- Identification follows using similar arguments.

Identification and estimation

A simple model

- For illustration, consider a special case of the 2-periods model.

- Job movers between ℓ and ℓ':

$$
\begin{aligned}
Y_{i1} &= a(\ell) + b(\ell)\alpha_i + \sigma(\ell)\varepsilon_{i1}, \\
Y_{i2} &= a(\ell') + b(\ell')\alpha_i + \sigma(\ell')\varepsilon_{i2},
\end{aligned}
$$

where α_i has mean $\mathbb{E}_{\ell\ell'}(\alpha_i)$ and variance $\text{Var}_{\ell\ell'}(\alpha_i)$, and ε_{it} are iid standard Gaussian.

- The ratio $b(\ell')/b(\ell)$ is **not** nonparametrically identified using mean and covariance restrictions.

- For example: not identified if α_i Gaussian (Reiersol, 1950).

A simple model (cont.)

- Now adding job movers between ℓ' and ℓ:

$$\begin{aligned} Y_{i1} &= a(\ell') + b(\ell')\alpha_i + \sigma(\ell')\varepsilon_{i1}, \\ Y_{i2} &= a(\ell) + b(\ell)\alpha_i + \sigma(\ell)\varepsilon_{i2}. \end{aligned}$$

- Then $b(\ell')/b(\ell)$ is **identified**, as:

$$\frac{b(\ell')}{b(\ell)} = \frac{\mathbb{E}_{\ell\ell'}(Y_{i2}) - \mathbb{E}_{\ell'\ell}(Y_{i1})}{\mathbb{E}_{\ell\ell'}(Y_{i1}) - \mathbb{E}_{\ell'\ell}(Y_{i2})},$$

provided that $b(\ell) \neq 0$, and that:

$$\mathbb{E}_{\ell\ell'}(\alpha_i) \neq \mathbb{E}_{\ell'\ell}(\alpha_i).$$

- **Empirical counterpart**: check whether

$$\mathbb{E}_{\ell\ell'}\left(Y_{i1} + Y_{i2}\right) \neq \mathbb{E}_{\ell'\ell}\left(Y_{i1} + Y_{i2}\right).$$

Identification of firm classes and worker composition

• To recover firm classes $f(j)$, note that, for all firms j such that $f(j) = \ell$ the cdf of log wages in period 1 is:

$$\Pr\left[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell\right] \;=\; \sum_{k=1}^{K} \pi_k(\ell) F_{k\ell}(y) \equiv H_\ell\left(y\right),$$

where H_ℓ, $\ell = 1, ..., L$, are univariate cdfs.

• Identifying the $f(j)$ thus amounts to solving a **classification problem**. This requires the H_ℓ to be distinct.

• Finally, given firm classes $f(j)$ and cdfs $F_{k\ell}$ it is immediate to recover the **type proportions** $\pi_k(\ell)$ from cross-sectional wages, through (CROSS).

Estimation 1st step: firm classes

• Let $\widehat{F}_j$ be the **empirical cdf of wages** in firm j, and n_j the number of workers in firm j.

• We estimate firm classes $f(j)$, for all firms $j \in \{1, ..., J\}$, by solving:

$$\min_{f(1),...,f(J),H_1,...,H_L} \sum_{j=1}^{J} n_j \sum_{d=1}^{D} \left(\widehat{F}_j(y_d) - H_{f(j)}(y_d) \right)^2,$$

where $y_1, ..., y_D$ is a grid of values.

• This is a **clustering** algorithm (weighted k-means).

• **Uniform consistency** of $\widehat{f}(j)$ can be shown by verifying the conditions of Theorem 2 in Bonhomme and Manresa (2015).

Estimation 2nd and 3rd steps: wage distributions and worker composition

• Given estimated firm classes we solve a **finite mixture** problem.

• Under parametric assumptions on $F_{k\ell}$ and $G_{k\ell}$, we maximize:

$$\sum_{i=1}^{N} \sum_{\ell=1}^{L} \sum_{\ell'=1}^{L} \mathbf{1}\{\widehat{f}(j_{i1}) = \ell\}\mathbf{1}\{\widehat{f}(j_{i2}) = \ell'\} \ln \left(\sum_{k=1}^{K} p_k(\ell,\ell') f_{k\ell}(Y_{i1};\theta) g_{k\ell'}(Y_{i2};\theta) \right).$$

• In the application we assume that $F_{k\ell}$ is Gaussian with (k,ℓ)-specific means and variances, and we use the EM algorithm.

• Lastly, given $\widehat{f}(j)$ and $\widehat{F}_{k\ell}$ we estimate $\pi_k(\ell)$ by maximizing:

$$\sum_{i=1}^{N} \sum_{\ell=1}^{L} \mathbf{1}\{\widehat{f}(j_{i1}) = \ell\} \ln \left(\sum_{k=1}^{K} \pi_k(\ell) f_{k\ell}(Y_{i1};\widehat{\theta}) \right).$$

Data and empirical results (preliminary)

Sample description

- We use four different databases covering the entire working age population in Sweden between 1997 and 2006.

- We follow Friedrich, Laun, Meghir and Pistaferri (2014, WP) for matching and sample selection.

- We select full-year employed **males** in 2002 (period 1). We refer to this sample of 800,000 workers and 50,000 firms as Sample 1.

Sample description (cont.)

• From Sample 1 we select job movers who are full-year employed in 2004 (period 2), and whose firm IDs are different in 2002 and 2004.

• In order to avoid considering job changes that are unrelated to job mobility, in the baseline we discard workers whose firm identifier is not present in 2002 or 2004.

• Sample 2 is the resulting sample of 20,000 **job movers**. [We also estimated the model on a sample of 55,000 movers without imposing the above selection.]

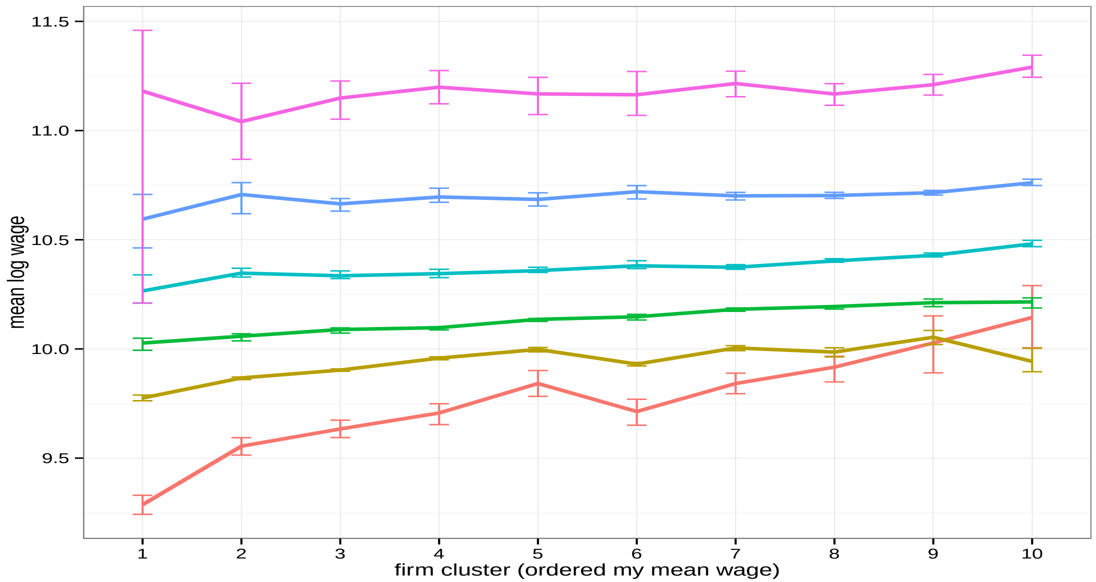
• In both samples we compute log pre-tax **annual earnings**, net of time dummies (within education*cohort).

Descriptive statistics on estimated firm classes

firm cluster:	1	2	3	4	5	6	7	8	9	10	all
number of workers	21,662	62,929	110,792	114,324	100,080	78,837	137,971	85,806	58,728	27,023	798,152
number of firms	6,487	7,972	7,804	6,494	4,663	3,748	4,209	3,984	3,157	2,812	51,330
% HS dropout	28.9	28	26.6	26.9	23.7	21.1	18.9	12.2	5.31	3	20.7
% HS grade	59.7	62.5	62.6	62.5	61.7	57.8	58.6	47.2	32.9	23.9	56.1
% some college	11.4	9.42	10.7	10.7	14.6	21.2	22.5	40.5	61.8	73.1	23.3
mean log wages	9.6	9.87	9.99	10.1	10.1	10.1	10.2	10.4	10.5	10.8	10.2
mean of log value added per worker	12.4	12.5	12.7	12.7	12.8	12.8	12.9	13	13	13.2	12.7

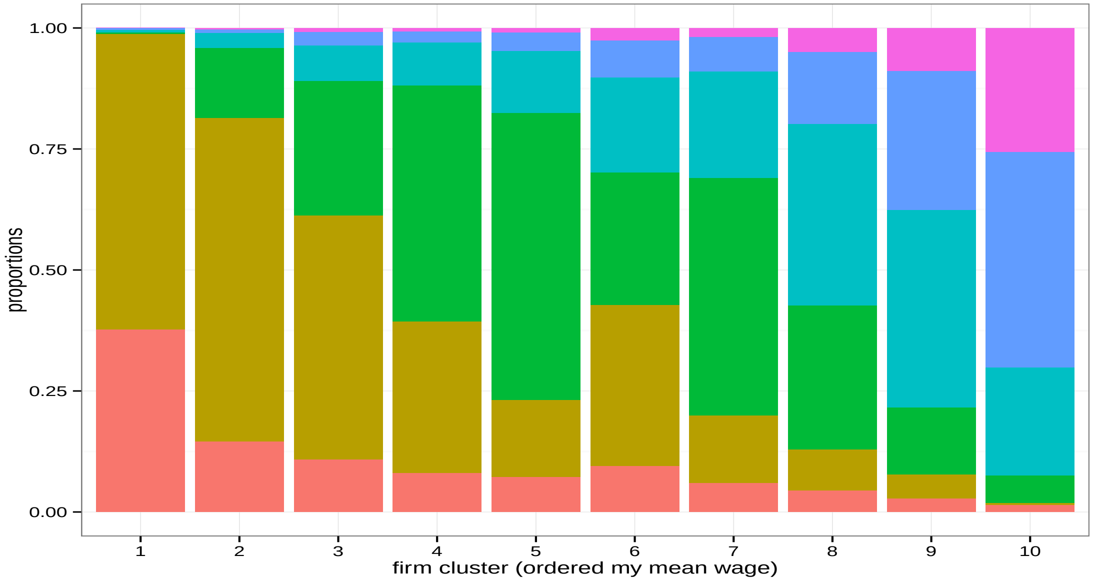
Notes: Sample 1 in 2002. All workers are males, employed during the full year 2002. "HS" is high school.

Estimated means of log wages by worker type and firm class (2-periods model)



Notes: The graph plots the mean of $\widehat{F}_{k\ell}$. The $L = 10$ firm classes (on the x-axis) are ordered by mean log wage. The $K = 6$ worker types correspond to the 6 different curves. 95% confidence intervals based on the parametric bootstrap (200 replications).

Estimated proportions of worker types by firm class (2-periods model)



Notes: Type proportions $\widehat{\pi}_k(\ell)$. *Firm classes are on the x-axis, type proportions on the y-axis. Colors correspond to the 6 different worker types.*

Simulations and decompositions

• We **simulate** the model based on the estimated parameters, conditional on the job moves in the data.

• We simulate entire employment spells, using the spell lengths in the data.

• We run **linear regressions** of the form:

$$Y_{i1} = \alpha[\omega^0(i)] + \psi[f^0(j_{i1})] + \varepsilon_{i1}$$

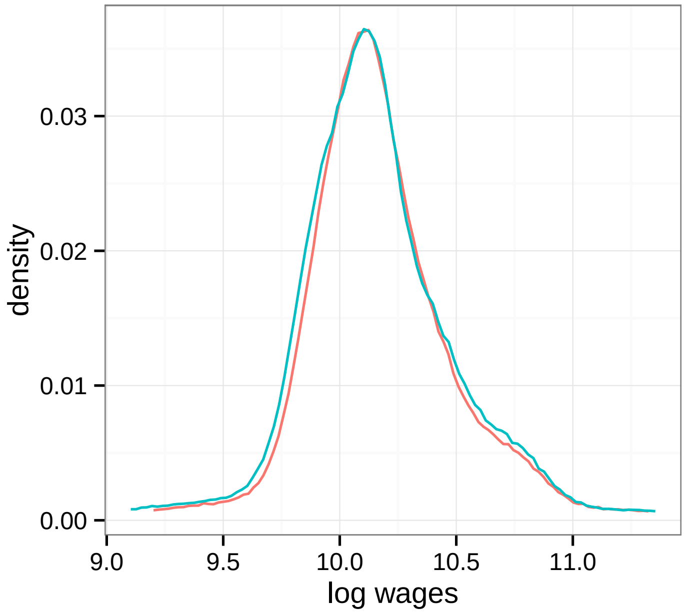
• We compare our results with fixed-effects regressions on real and simulated data.

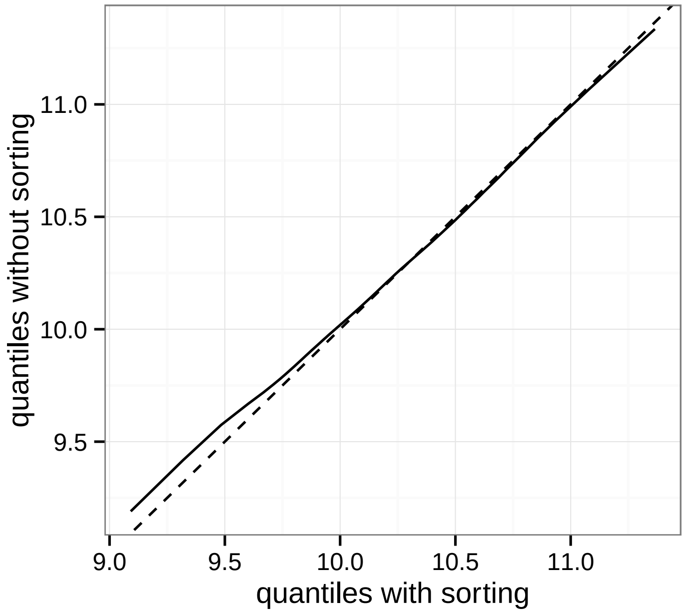
Variance decompositions on Swedish data and simulated data (2-periods model)

	min spell	rep	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$
			Data			
This paper			0.7766	0.0473	0.1762	0.4598
Fixed-effects			0.9813	0.3014	-0.2826	-0.2599
			Simulated from the model			
This paper	1	1	0.7669	0.0466	0.1866	0.4934
Fixed-effects	1	1	1.0879	0.3447	-0.4326	-0.3532
			Simulated from the model without limited mobility			
Fixed-effects	4	1	0.8948	0.1602	-0.055	-0.0727
Fixed-effects	4	10	0.7816	0.053	0.1654	0.4064

Notes: Real and simulated data. α is the worker (or type) fixed-effect, ψ is the firm (or class) fixed-effect. "min spell" is the minimum length of employment spells. "rep" is the number of job movers per firm, relative to the original dataset.

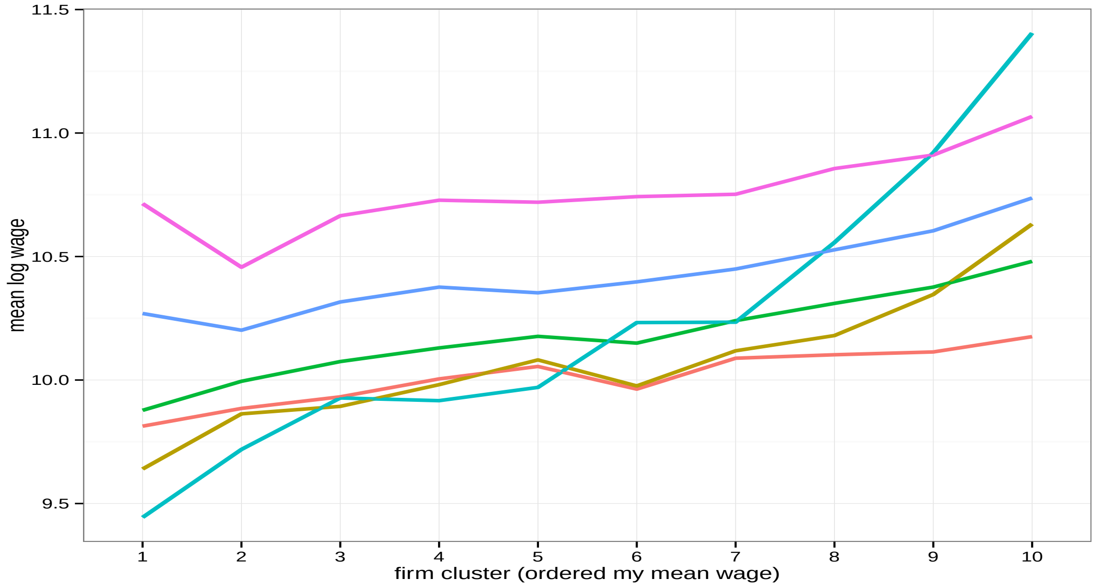
Wage distribution with/without sorting





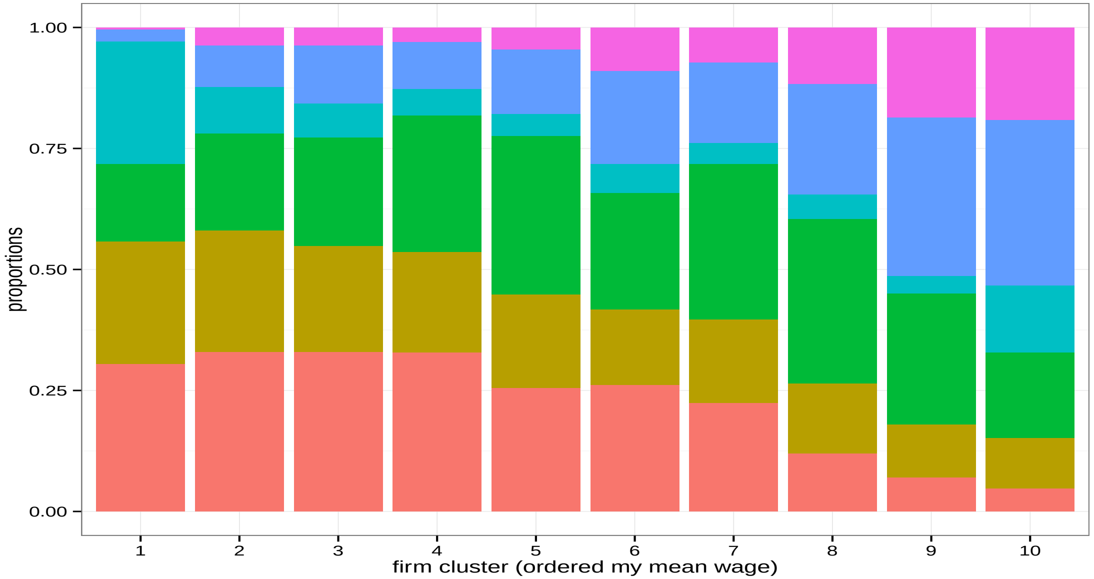
Notes: The left graph plots the log wage densities, Sample 1, 2002. Blue is simulated from the model, red is simulated from the model without sorting. The right graph plots quantiles of log wages. The log wage variance without sorting is .127, versus .144 with sorting (88%).

Estimated means of log wages by worker type and firm class (4-periods model, preliminary)



Notes: The graph plots the mean of $\widehat{F}_{k\ell}$ *for 2002 in the 4-periods model. Firm classes are on the x-axis. The* 6 *worker types correspond to the different curves.*

Estimated proportions of worker types by firm class (4-periods model, preliminary)



Notes: Type proportions $\widehat{\pi}_k(\ell)$ in the 4-periods model. Firm classes are on the x-axis, type proportions on the y-axis. Colors correspond to the 6 different worker types.

Variance explained by workers, firms, and covariance: 50% (60% when taking out the steepest type), 29% (20%), and 21% (20%). Correlation: 28% (29%).

Conclusion

- Econometric framework for matched data.

- The clustering approach allows to get back to standard single-agent econometric models.

- Application to structural models where joint estimation of the heterogeneity and the full model might be computationally prohibitive.

- Beyond workers and firms: sorting across schools, neighborhoods, cities...

Additional slides

Job mobility (2-periods model, cont.)

• Let $m_i = 1\{j_{i1} \neq j_{i2}\}$ denote job mobility. Two assumptions:

i) **Mobility driven by types/classes**: job movements can depend on worker type and firm classes (before and after the move), but not on wages.

$$\begin{aligned}&\Pr\left[m_i = 1, f(j_{i2}) = \ell' \,|\, \omega(i) = k, f(j_{i1}) = \ell, \boldsymbol{Y_{i1}} = \boldsymbol{y}\right] \\ &\quad = \Pr\left[m_i = 1, f(j_{i2}) = \ell' \,|\, \omega(i) = k, f(j_{i1}) = \ell\right].\end{aligned}$$

ii) **Serial independence**: wages drawn from different firms are independent, conditional on worker type and firm class.

$$\begin{aligned}&\Pr\left[Y_{i2} \leq y' \,|\, m_i = 1, \omega(i) = k, f(j_{i2}) = \ell', \boldsymbol{f(j_{i1})} = \boldsymbol{\ell}, \boldsymbol{Y_{i1}} = \boldsymbol{y}\right] \\ &\quad = \Pr\left[Y_{i2} \leq y' \,|\, m_i = 1, \omega(i) = k, f(j_{i2}) = \ell'\right].\end{aligned}$$

Job mobility (4-periods model, cont.)

- Let $m_i = \mathbf{1}\{j_{i2} \neq j_{i3}\}$. Two assumptions:

$$i) \quad \Pr\left[m_i = 1, f(j_{i3}) = \ell', Y_{i3} \leq y_3 \,|\, \omega(i) = k, f(j_{i2}) = \ell, Y_{i2} = y_2, \boldsymbol{Y_{i1}} = \boldsymbol{y_1}\right]$$
$$= \Pr\left[m_i = 1, f(j_{i3}) = \ell', Y_{i3} \leq y_3 \,|\, \omega(i) = k, f(j_{i2}) = \ell, Y_{i2} = y_2\right].$$

$$ii) \quad \Pr\left[Y_{i4} \leq y_4 \,|\, m_i = 1, \omega(i) = k, f(j_{i3}) = \ell', Y_{i3} = y_3, \boldsymbol{f(j_{i2})} = \boldsymbol{\ell}, \boldsymbol{Y_{i2}} = \boldsymbol{y_2}, \boldsymbol{Y_{i1}} = \boldsymbol{y_1}\right]$$
$$= \Pr\left[Y_{i4} \leq y_4 \,|\, m_i = 1, \omega(i) = k, f(j_{i3}) = \ell', Y_{i3} = y_3\right].$$

Incorporating covariates

• Let X_{it} and B_{jt} denote covariates. The DGP in period 1 is as follows:

-For all j, B_{j1} is a draw from an ℓ-specific distribution.

-For all i, X_{i1} is a draw from an (ℓ, b)-specific distribution.

-For all i, $\omega(i)$ is a draw from an (ℓ, b, x)-specific distribution.

-For all i, Y_{i1} is a draw from an (k, ℓ, b, x)-specific distribution.

• The $f(j)$ can then be recovered by clustering firm wage distributions. Alternatively, one can cluster firm distributions of $(Y_{i1}, X_{i1}, B_{j(i,1),1})$.

• Then, covariates may be added to the wage distributions and type proportions ($F_{kx\ell b}$, $p_{kx_1x_2}(\ell, b_1, \ell', b_2)$, $\pi_{kx}(\ell, b)$) in the second and third estimation steps.

Link to theoretical models

- 2-periods model:

-Example: Shimer and Smith (2000), without or with on-the-job search (workers' threat points being the value of unemployment).
-No role for match-specific draws, unless independent over time or measurement error. No sequential auctions.

- 4-periods model:

-All models where state variables (k, ℓ_t, Y_t) are **first-order Markov**.
-Examples: wage posting, sequential auctions (Lamadon, Lise, Meghir and Robin 2015).
-No latent human capital accumulation (k_t), no permanent+transitory within-job wage dynamics (example: random walk+i.i.d. shock).

Identification of wage distributions

• More generally, let $\ell \neq \ell'$ be firm classes, and write (MOV) in **matrix form** for (ℓ, ℓ') transitions:

$$A(\ell, \ell') = F(\ell)D(\ell, \ell')G(\ell')'.$$

• The identification argument relies on the **joint diagonalization** of a set of matrices.

• In the stationary case (i.e., $F_{k\ell} = G_{k\ell}$) it is sufficient (but not necessary) that:

-For all ℓ there exists an ℓ' such that $p_k(\ell, \ell') \neq 0$ for all k, and the ratios $p_k(\ell', \ell)/p_k(\ell, \ell')$, $k = 1, ..., K$, are distinct.

-The columns of $F(\ell)$ (that is, the cdfs $F_{k\ell}$) are linearly independent.

Identification of wage distributions (cont.)

• Consider for simplicity the case where $F_{k\ell} = G_{k\ell}$ (i.e., stationarity). Then a singular value decomposition of $A(\ell, \ell')$ yields:

$$A(\ell, \ell') = F(\ell)D(\ell, \ell')F(\ell')' = USV',$$

where U and V are orthogonal with K columns, and S is diagonal.

• Letting $W_1 = S^{-\frac{1}{2}}U'$ and $W_2 = S^{-\frac{1}{2}}V'$, we have:

$$W_1 A(\ell', \ell)' W_2' = W_1 F(\ell) D(\ell, \ell')^{-1} D(\ell', \ell) D(\ell, \ell') F(\ell')' W_2',$$

where $D(\ell, \ell')F(\ell')'W_2' = (W_1 F(\ell))^{-1}$.

• Hence $W_1 F(\ell)$ is a matrix of eigenvectors, the corresponding eigenvalues being the diagonal elements of $D(\ell, \ell')^{-1} D(\ell', \ell)$.

• Identification of $F(\ell)$ then follows from the fact that $F_{k\ell}$ are cdfs.

Identification of firm classes (cont.)

• When adding information from **job movers**, conditions needed for identification are weaker:

$$\text{Pr}\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] \;\equiv\; G_{\ell\ell'}\left(y, y'\right).$$

• Identification of $f(j)$ requires that for all $\ell \neq \ell'$ there exists an ℓ'' such that $G_{\ell\ell''} \neq G_{\ell'\ell''}$ or $G_{\ell''\ell} \neq G_{\ell''\ell'}$.

• In this talk we cluster firms' cross-sectional distributions.

Identification of wage distributions: an example

• Suppose $K = L$, and that types and classes are ordered. Suppose that $\pi_k(\ell) \neq 0$ iff $|k - \ell| \leq 1$, and that $p_k(\ell, \ell') \neq 0$ iff $(|k - \ell| \leq 1, |k - \ell'| \leq 1)$.

• Then $\operatorname{rank} A(1,2) = \operatorname{rank} A(2,1) = 2$. It follows as in the main analysis that (F_{11}, F_{21}) and (F_{12}, F_{22}) are identified.

• Likewise, $\operatorname{rank} A(2,3) = \operatorname{rank} A(3,2) = 3$. It follows that, for some (k_1, k_2, k_3), $(F_{k_1 2}, F_{k_2 2}, F_{k_3 2})$ and $(F_{k_1 3}, F_{k_2 3}, F_{k_3 3})$ are identified.

• As $F(2)$ has full column rank, we can identify which of the (k_1, k_2, k_3) are equal to 1 or 2. Without loss of generality, let $k_1 = 1$ and $k_2 = 2$. Set $k_3 = 3$. Then (F_{12}, F_{22}, F_{32}) and (F_{13}, F_{23}, F_{33}) are identified.

Identification of wage distributions: an example (cont.)

• Continuing the argument we identify:

-(F_{11}, F_{21})
-(F_{12}, F_{22}, F_{32})
-...
-$(F_{L-2,L-1}, F_{L-1,L-1}, F_{L,L-1})$
-$(F_{L-1,L}, F_{LL})$

• The other $F_{k\ell}$'s are not identified. These correspond to the (k, ℓ) combinations such that $\pi_k(\ell) = 0$.

• In this example, without additional structure one cannot assess the wage effects of randomly allocating workers to jobs.

Theoretical search-matching model: setup

• Worker x and firm y, on-the-job search (λ_0, λ_1).

-Firms post vacancies.

-Production function $f(x,y) = a + (\nu x^\rho + (1-\nu)y^\rho)^{1/\rho}$.

• Surplus equation is given by:

$$(r+\delta)S(x,y) = (1+r)\left(f(x,y) - \delta(b(x) - c(y))\right) - r(1-\delta)(\Pi_0(y) + W_0(x)) \\ + (1-\delta)\lambda_1 \int P(x,y,y')(\alpha S(x,y') - S(x,y))v_{1/2}(y')dy',$$

where $P(x,y,y')$ is 1 when $S(x,y) > S(x,y')$.

• Wage equation is given by:

$$(1+r)w(x,y) = (r+\delta)\alpha S(x,y) + (1-\delta)rW_0(x) - \\ (1-\delta)\lambda_1 \int P(x,y,y')(\alpha S(x,y') - \alpha S(x,y))v_{1/2}(y')dy'.$$

Institutional Background in Sweden

- The literature has documented between-firm/plant wage variation in Sweden to be lower compared to other countries such as Germany or Brazil (e.g. Akerman *et al.* 2013, and Baumgarten *et al.* 2014).

- Potential factors: highly unionized labor market, and tradition of collective wage bargaining agreements in Sweden.

- In particular, after World War II the introduction of the so-called solidarity wage policy, which had as guiding principle "equal pay for equal work", significantly limited the capacity of firms to differentially reward their employees.

- However, several reforms since the 1990s have contributed to an increase in between-firm wage variation due to a more informal coordination in wage setting. See Skans *et al.* (2009) for more details.

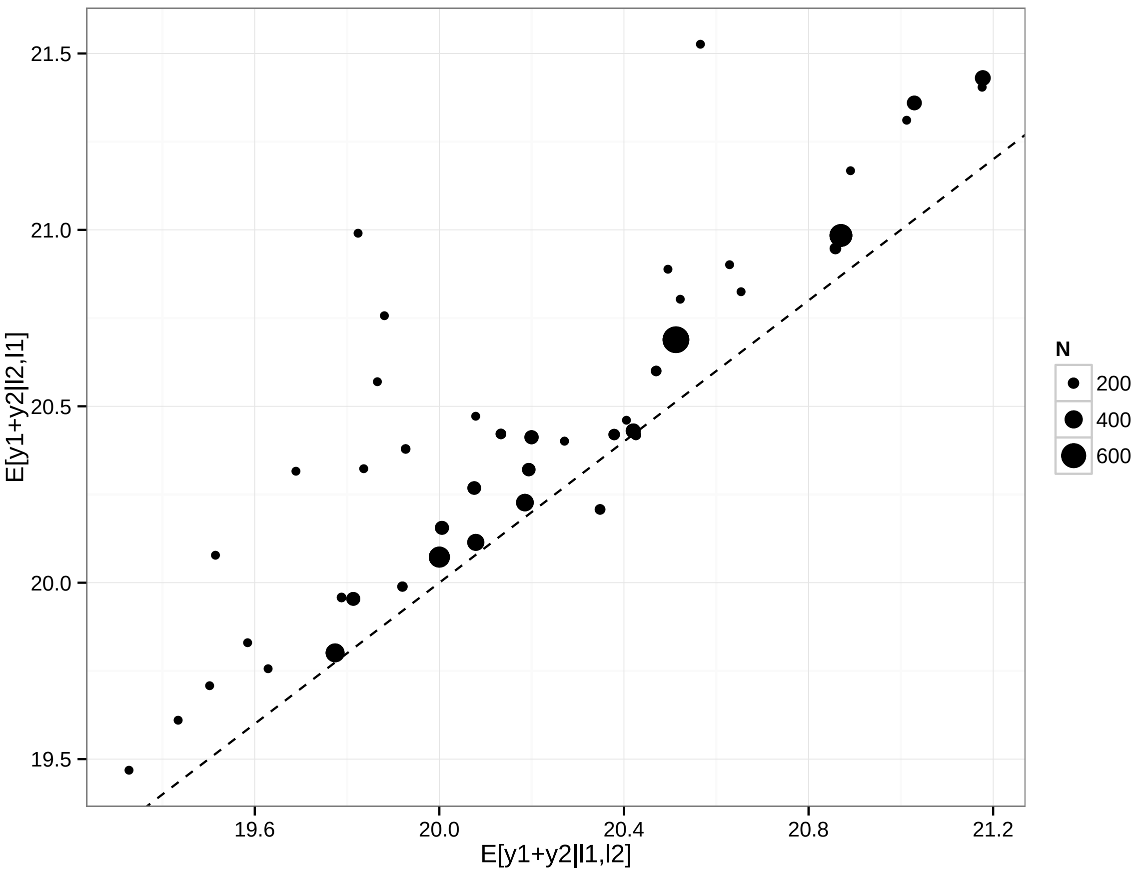
Descriptive statistics and data selection

	all	employed either	employed both	continuing firms
firms in 2002	54,753	53,610	46,597	43,884
firms in 2003	55,623	54,674	47,553	43,845
firms in 2004	56,374	54,867	46,450	43,887
workers in 2002	1,091,509	907,883	635,186	599,963
workers in 2003	1,082,028	910,454	635,135	598,834
workers in 2004	1,073,174	886,573	635,186	599,963
mean reported firm size in 2002	34.4	34.9	37.6	37.1
median reported firm size in 2002	10	10	10	10
movers between 2002 and 2004	142,580	121,090	54,968	19,745
% movers employed 12 months in 2003	0.755	0.868	0.952	0.968
co-movers 90 percentile	2	2	2	1
co-movers 99 percentile	13	15	22	7
co-movers 100 percentile	2,458	2,439	2,137	233
quaterly j2j probability	0.0196	0.0183	0.0149	0.00678
quaterly e2u probability	0.0228	0.0133	0.00396	0
quaterly u2e probability	0.155	0.34	0.477	-

Notes: Description of the data in different samples. The movers in the rightmost column correspond to Sample 2. Note that Sample 1, which contains all workers full-year employed in 2002, is not in this table.

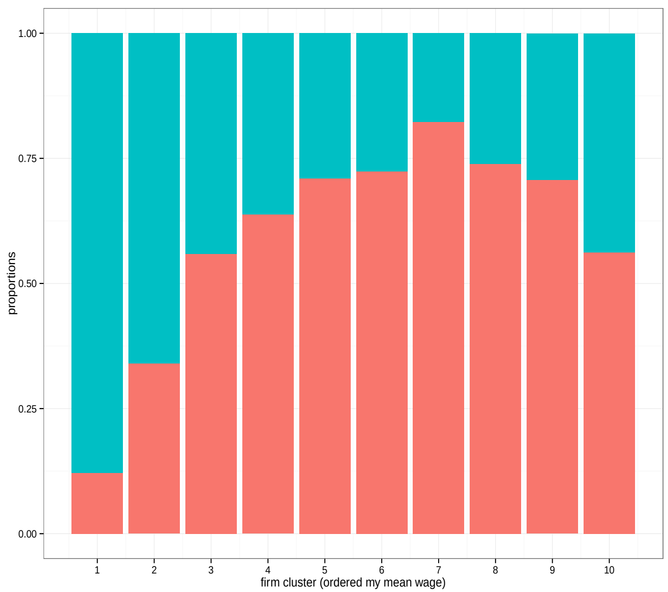
Wages of job movers

$\mathbb{E}_{\ell_1\ell_2}(Y_{i1}+Y_{i2})$ (x-axis) vs $\mathbb{E}_{\ell_2\ell_1}(Y_{i1}+Y_{i2})$ (y-axis), $\ell_1 < \ell_2$

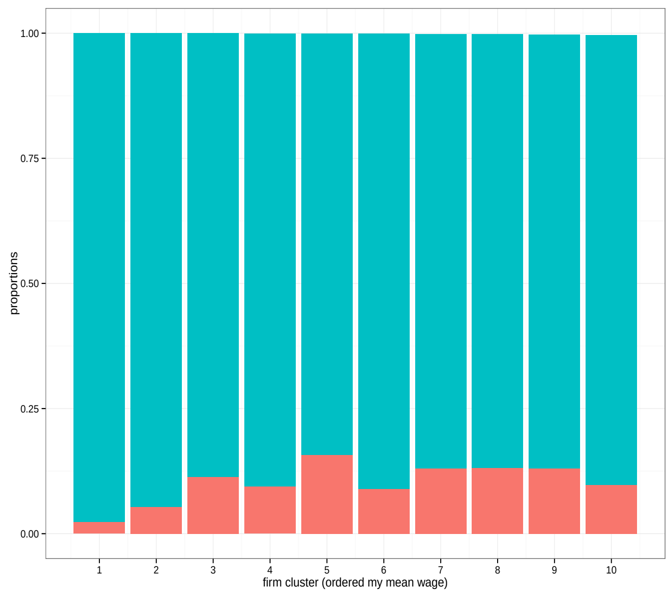


Proportions of firms/workers in large (≥ 50) and small (< 50) firms, by firm class

Proportions of workers



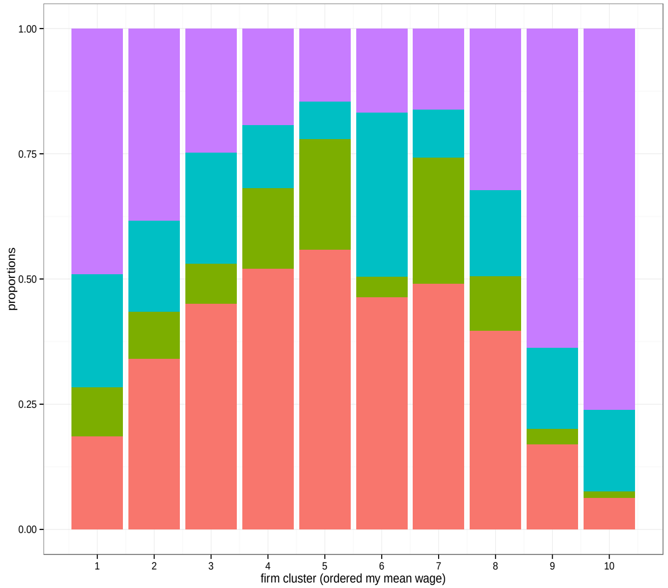
Proportions of firms



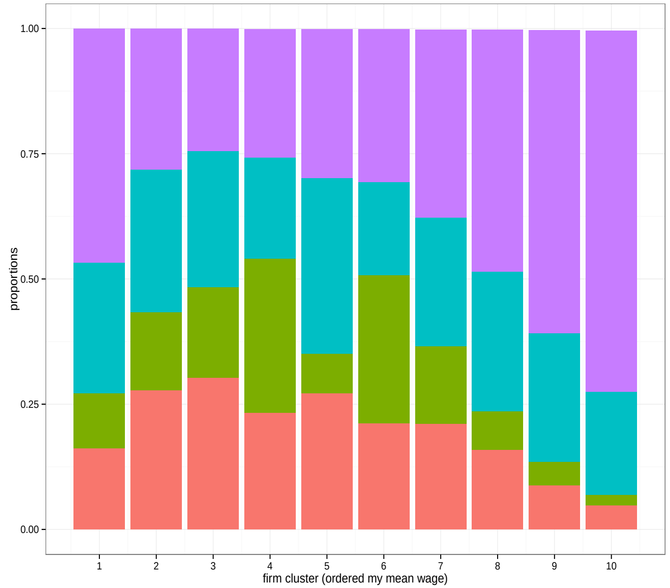
Notes: Blue (small) and red (large).

Proportions of workers in different industries, by firm class

Proportions of workers



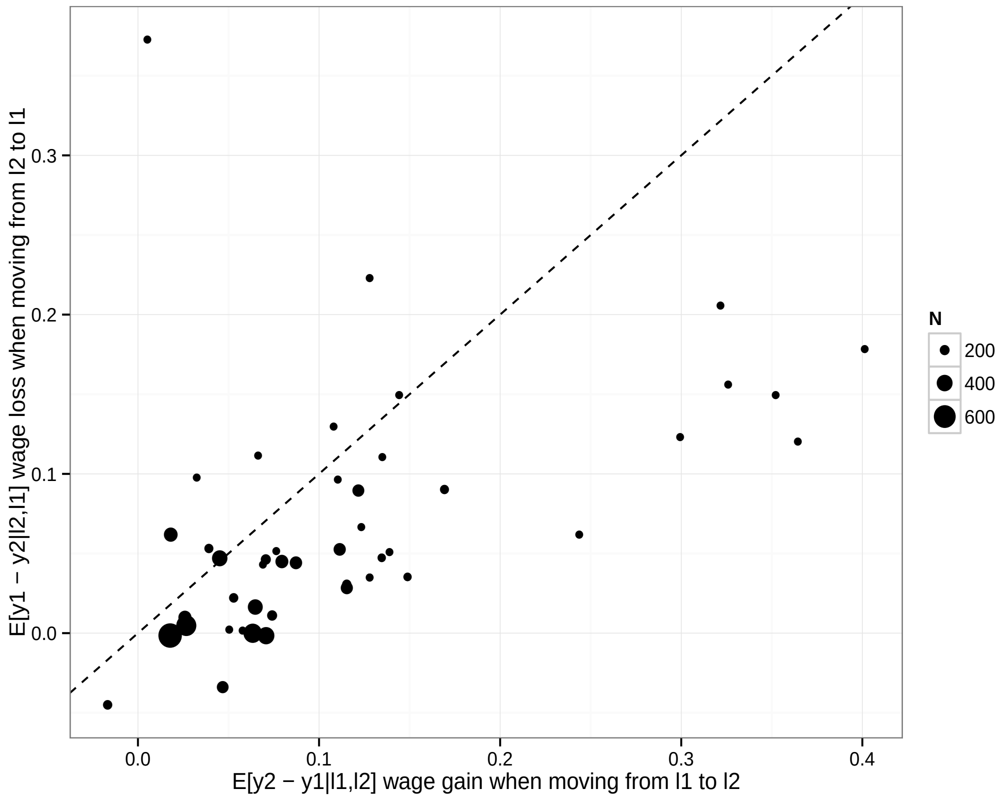
Proportions of firms



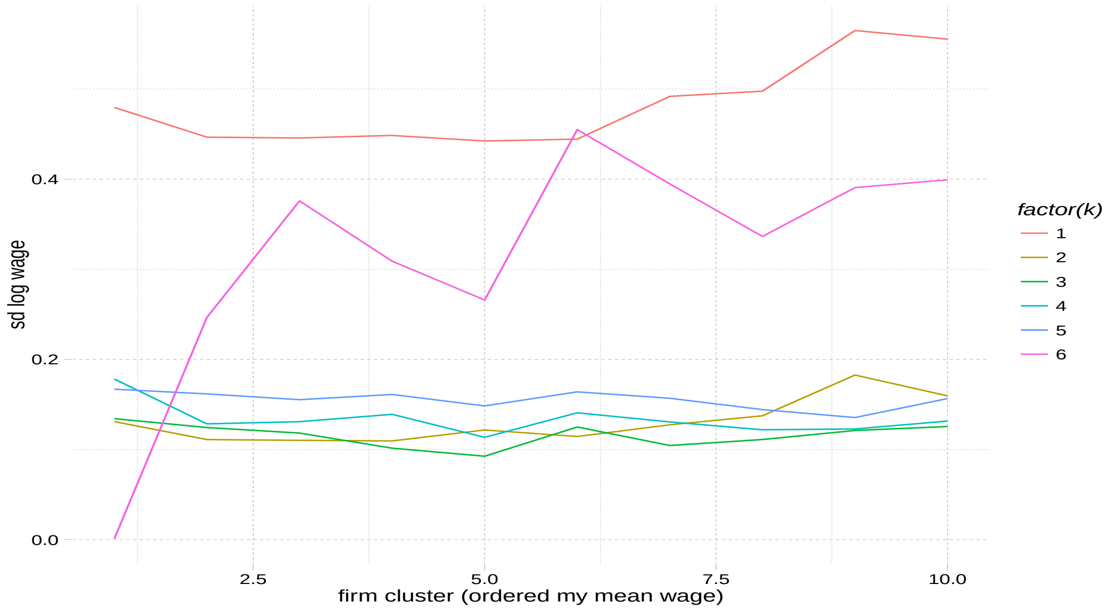
Notes: Red (manufacturing), green (construction), blue (retail), and purple (services).

Wages of job movers

$\mathbb{E}_{\ell_1\ell_2}(Y_{i2} - Y_{i1})$ **(x-axis) vs** $\mathbb{E}_{\ell_2\ell_1}(Y_{i1} - Y_{i2})$ **(y-axis),** $\ell_1 < \ell_2$

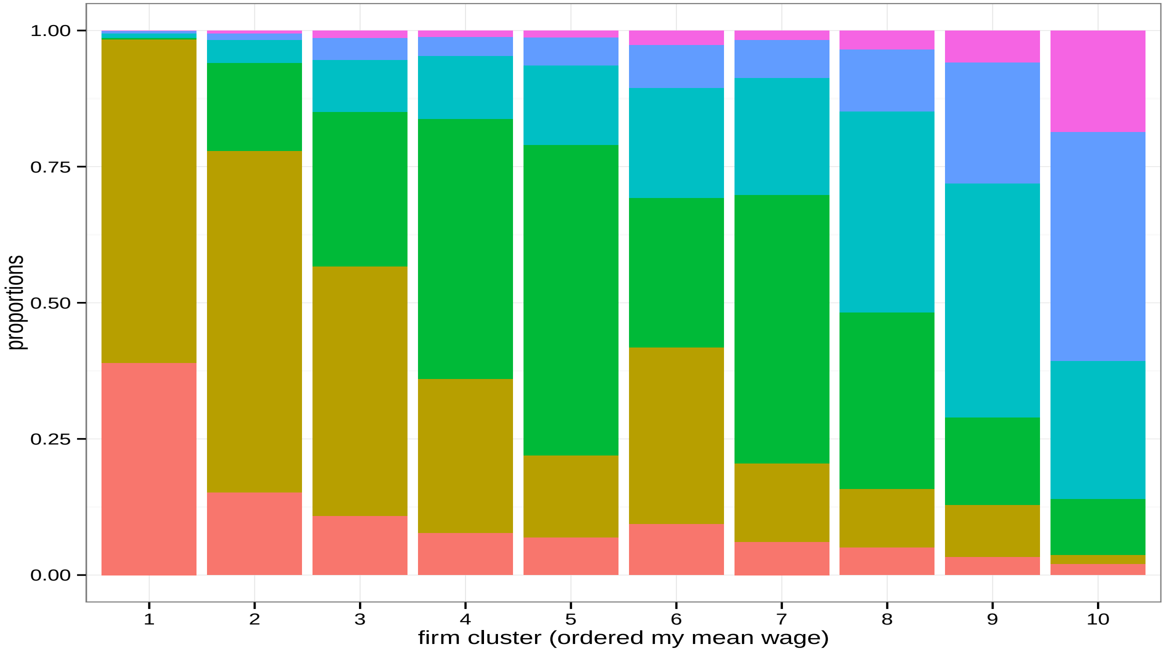


Estimated standard deviations of log wages by worker type and firm class (2-periods model)



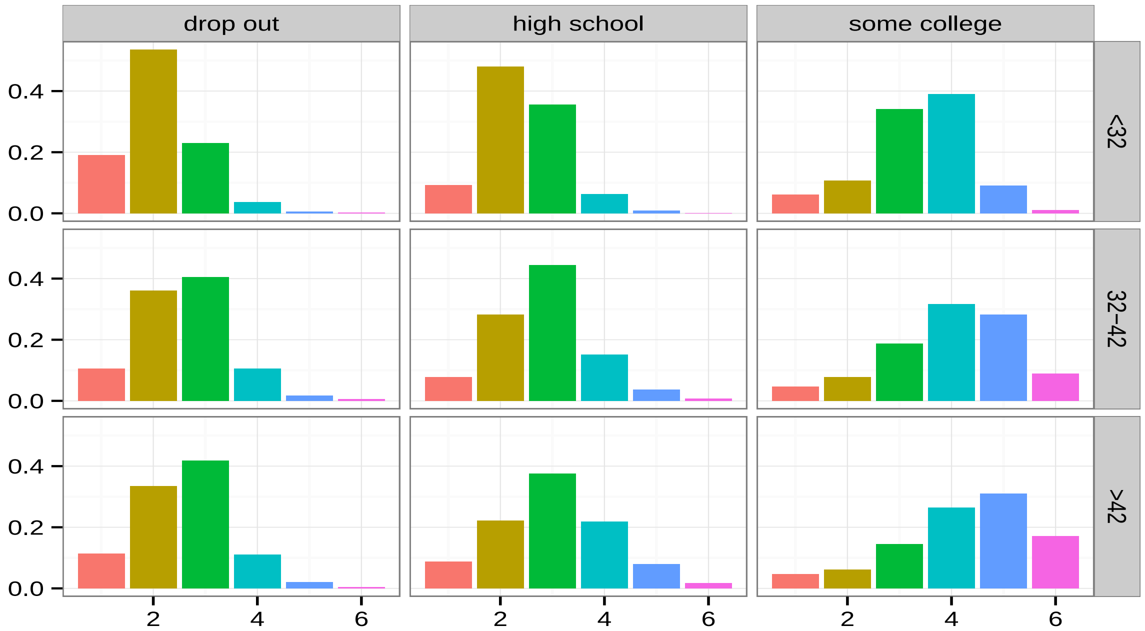
Notes: The graph plots the standard deviation of $\widehat{F}_{k\ell}$. The $L = 10$ firm classes (on the x-axis) are ordered by mean log wage. The $K = 6$ worker types correspond to the 6 different curves.

Estimated proportions of worker types by firm class (2-periods model, no covariates)



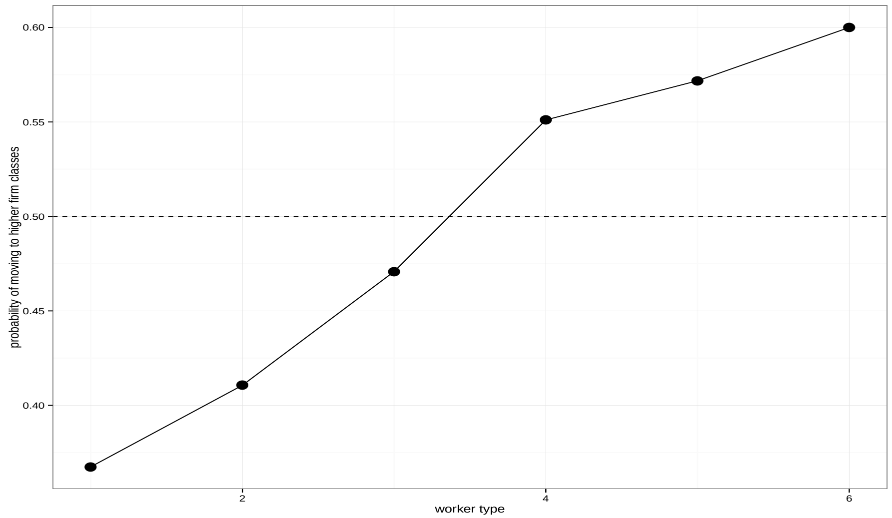
Notes: Type proportions net of differences in education and age x: $\widetilde{\pi}_k(\ell) = \sum_x \widehat{p}_x \widehat{\pi}_k(\ell, x)$*, where* $\widehat{p}_x$ *is the number of workers in cell* x*. Each plot corresponds a different firm class. Average log wages by* (k, ℓ) *combinations are on the x-axis, proportions are on the y-axis.*

Estimated proportions of worker types by education and age categories (2-periods models)



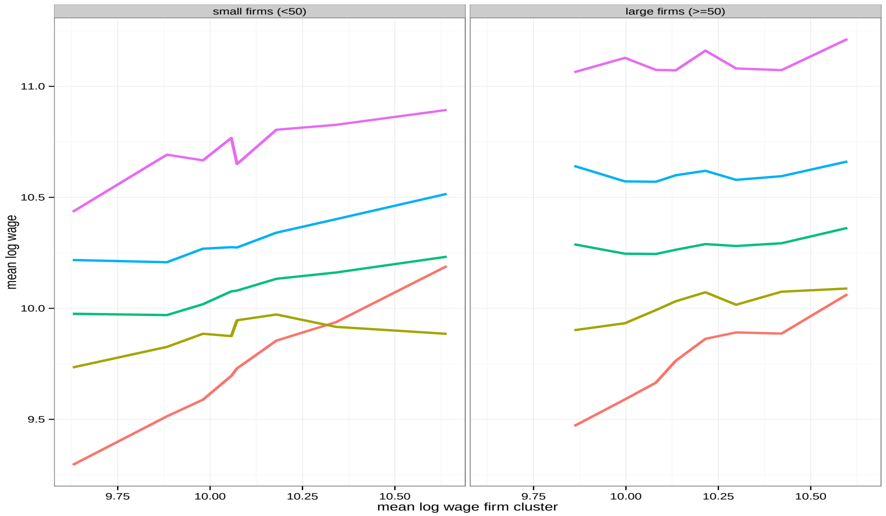
Notes: Type proportions by education and age categories.

Probability of moving to a higher firm class (2-periods model)



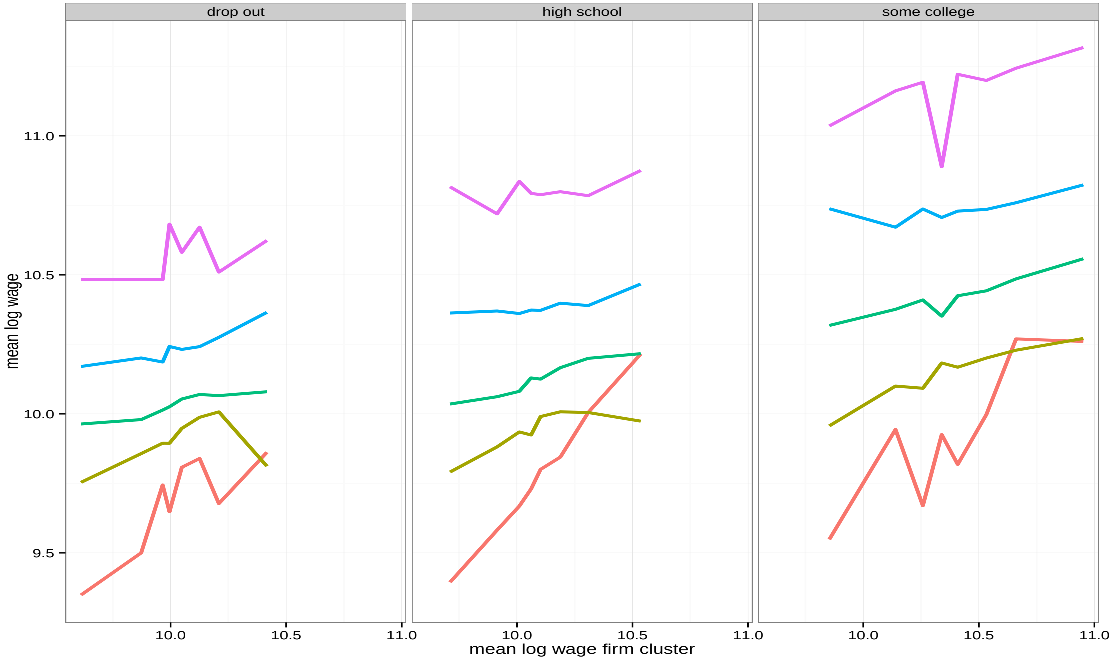
Notes: This graph plots the estimated probability of moving to a higher firm class (y-axis), by worker type (x-axis), averaged over classes.

Estimated means of log wages by worker type and firm class, by firm size (2-periods model)



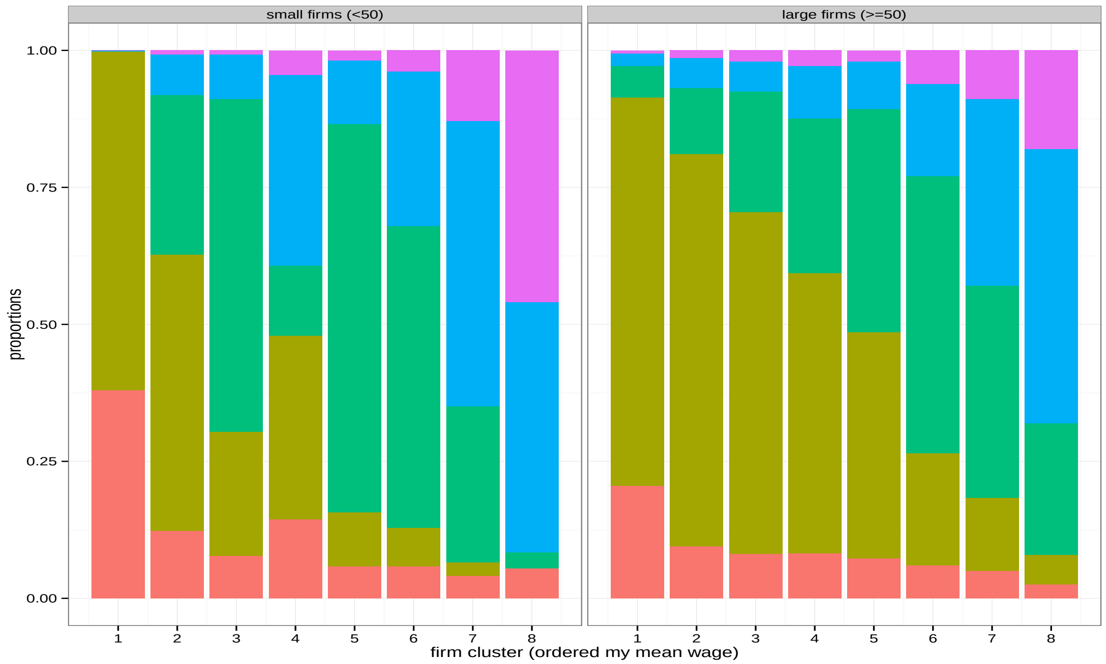
Notes: The graph plots the mean of $\widehat{F}_{k\ell}$, for small and large firms. The wages in the 8 firm classes are shown on the x-axis. The 5 worker types correspond to the different curves.

Estimated means of log wages by worker type and firm class, by worker education (2-periods model)



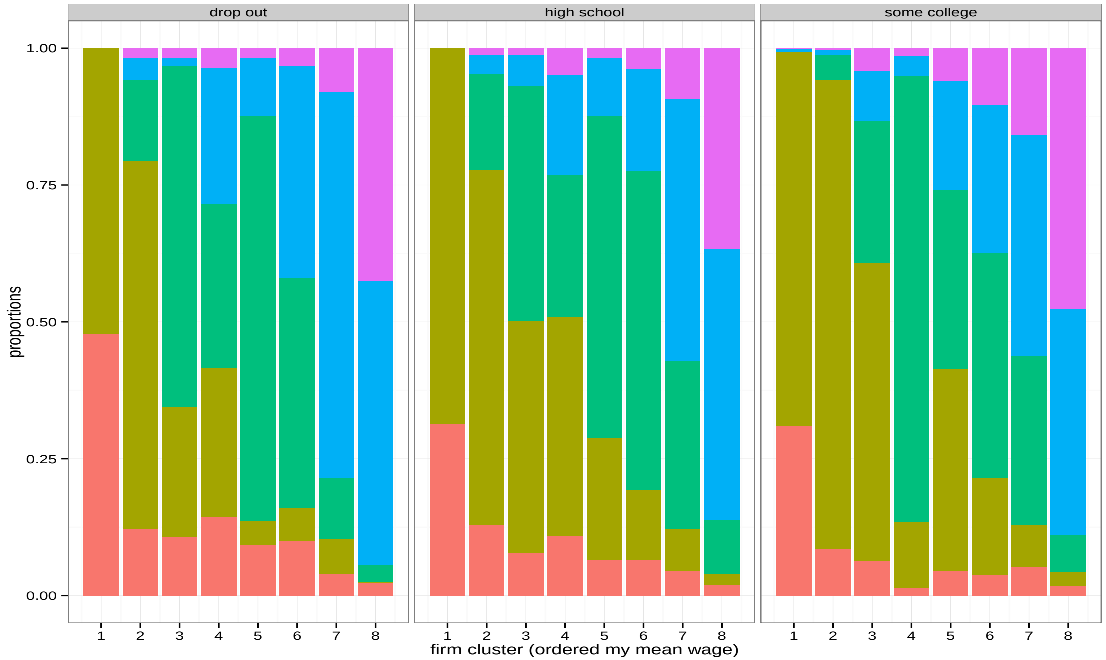
Notes: The graph plots the mean of $\widehat{F}_{k\ell}$, for high school drop-outs, high-school graduates, and some college. The wages in the 8 firm classes are shown on the x-axis. The 5 worker types correspond to the different curves.

Estimated proportions of worker types by firm class, by firm size (2-periods model)



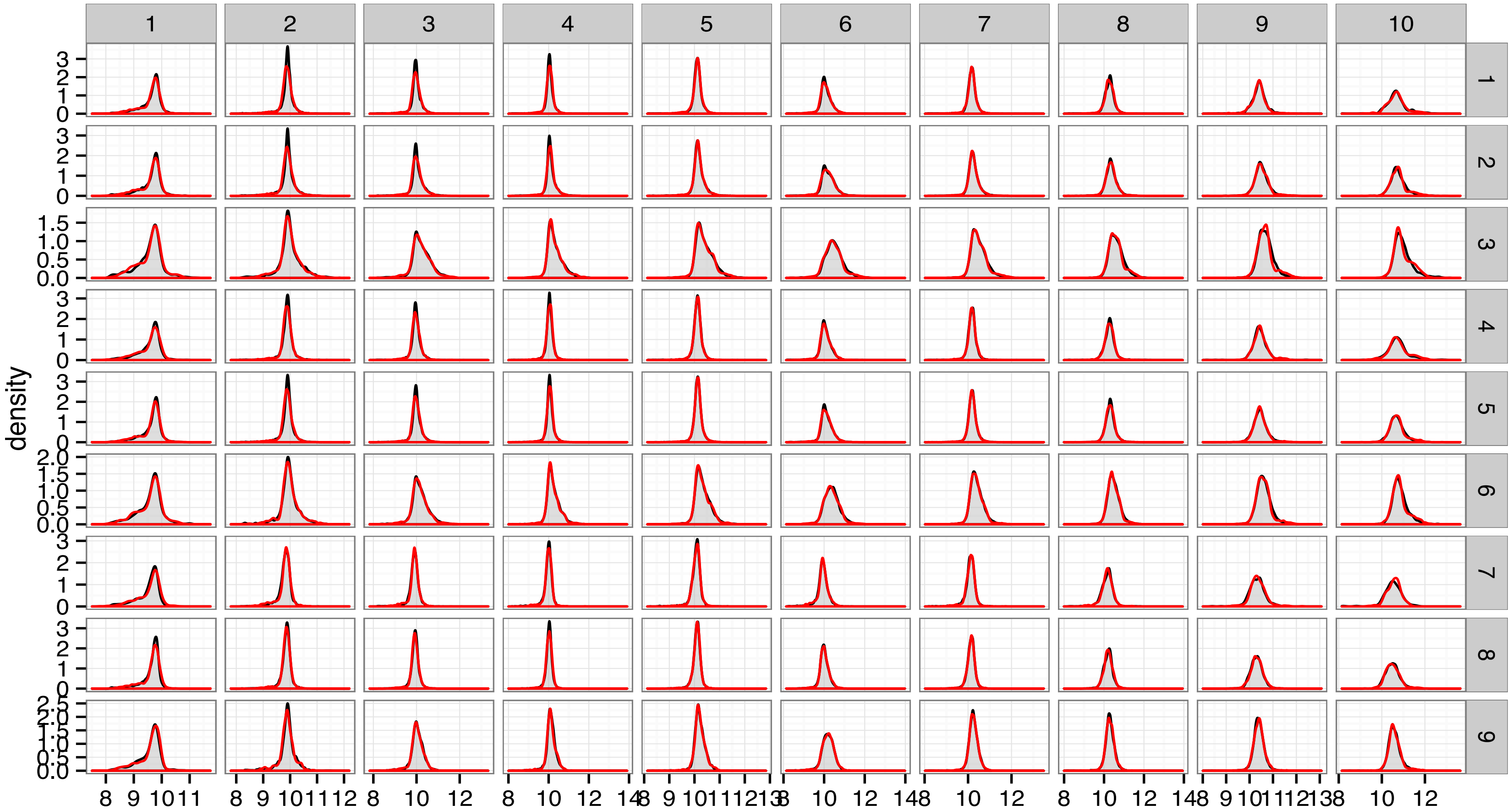
Notes: Small (< 50) and large (≥ 50) firms.

Estimated proportions of worker types by firm class, by worker education (2-periods model)



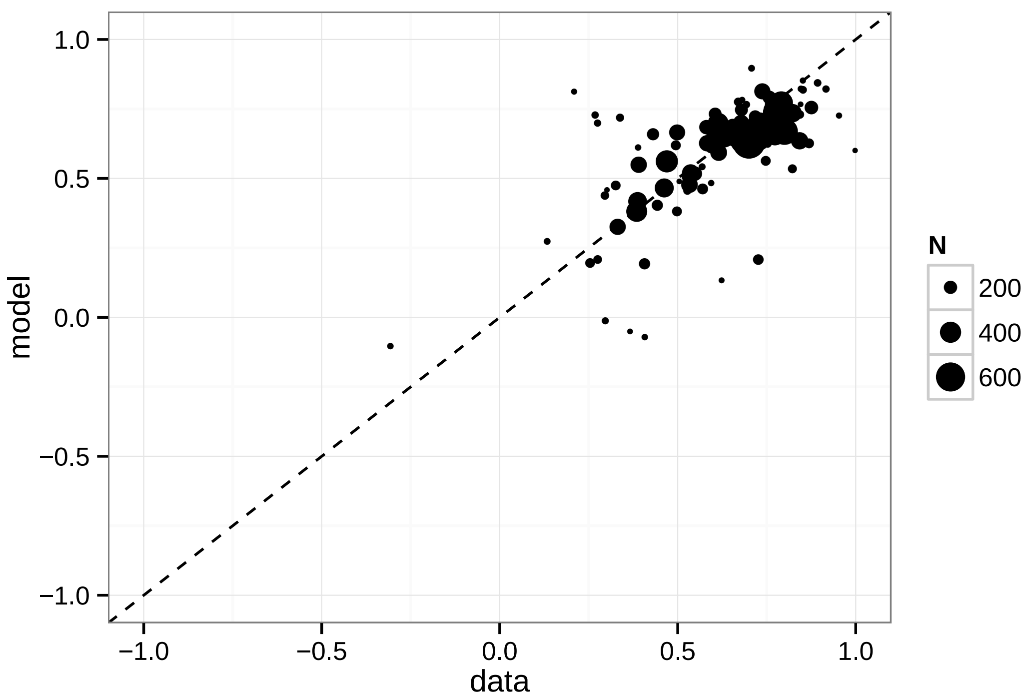
Notes: High school drop-outs, high-school graduates, and some college.

Fit of log wage densities (2-periods model)



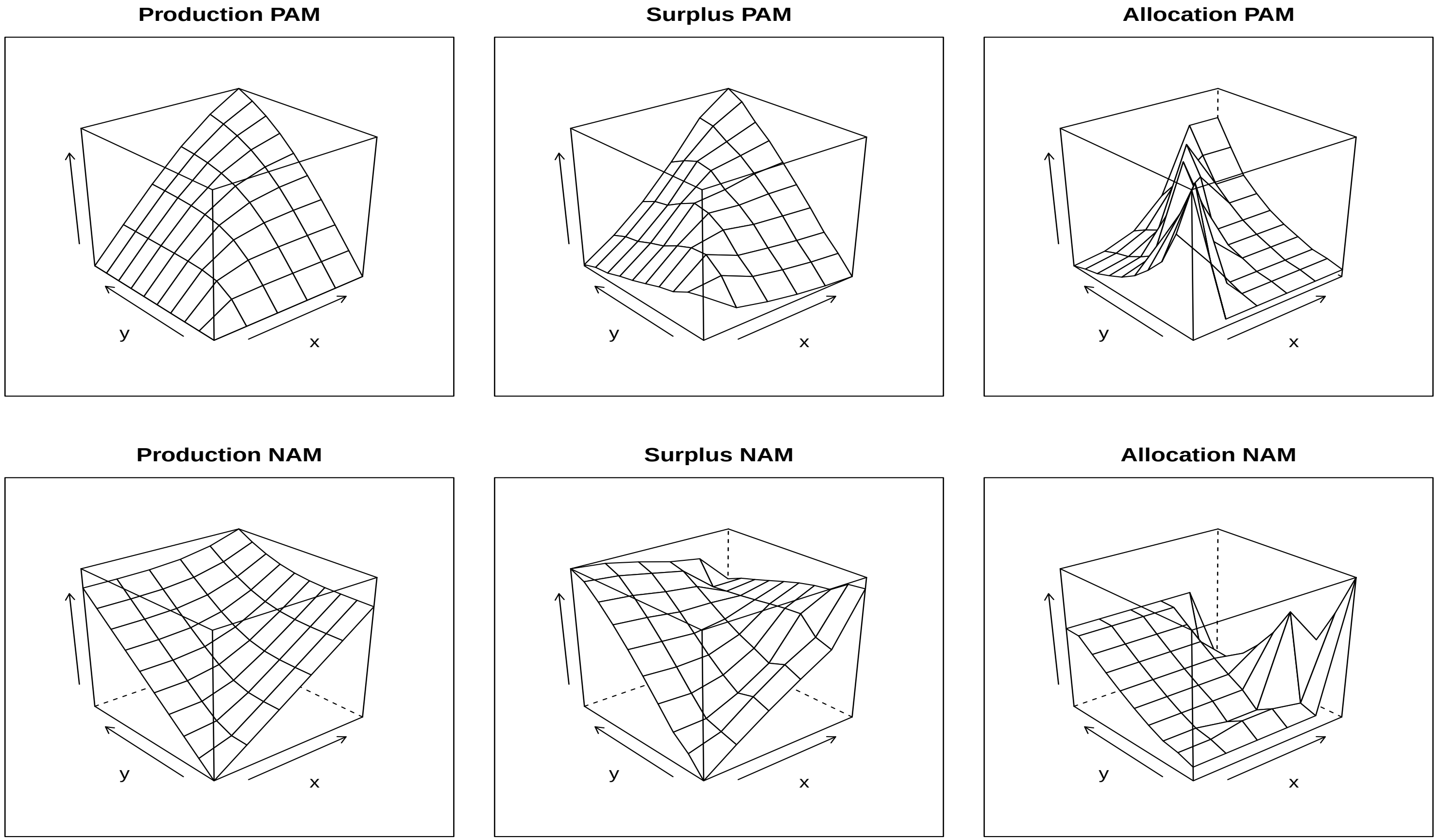
Notes: Marginal densities of log wages for each x cell (in rows) and firm class (in columns). Sample 1, 2002. The red line is the model, the shaded area is from the data.

Fit of log wage correlations (2-periods model)



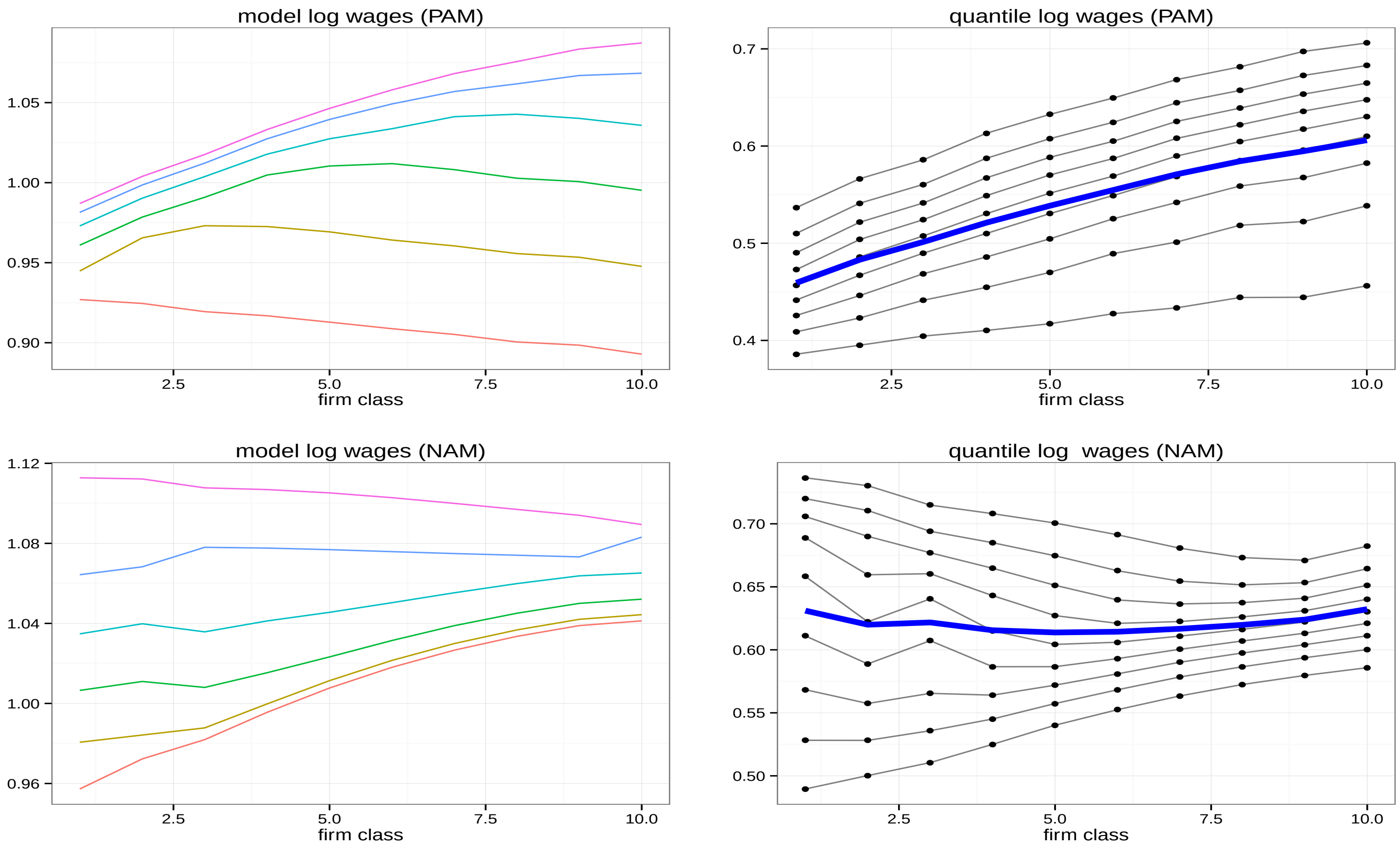
Notes: Log wage correlations $Corr(Y_1, Y_2|\ell_1, \ell_2)$*, for job movers, by pairs of firm classes. Sample 2. In the data (x-axis) and in the simulated data (y-axis).*

Theoretical search-matching model: setup



Notes: Model based on Shimer and Smith (2000) with on-the-job search.

Theoretical search-matching model: wage distributions



Notes: Model based on Shimer and Smith (2000) with on-the-job search.

Theoretical search-matching model: simulation results

		dim	%bw	%wwbf	%wwwf	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$
PAM	model	6×10	0.693	0.103	0.203	0.791	0.054	0.156	0.377
	BLM	6×10	0.636	0.101	0.263	0.756	0.069	0.175	0.385
NAM	model	6×10	0.661	0.136	0.203	1.082	0.125	-0.206	-0.281
	BLM	6×10	0.625	0.114	0.262	1.049	0.099	-0.148	-0.23
PAM	model	50×50	0.693	0.108	0.2	0.758	0.071	0.171	0.367
	BLM	6×10	0.591	0.121	0.288	0.701	0.095	0.204	0.396
NAM	model	50×50	0.685	0.115	0.201	1.079	0.107	-0.186	-0.273
	BLM	6×10	0.668	0.044	0.288	1.009	0.041	-0.05	-0.122

Notes: Model based on Shimer and Smith (2000) with on-the-job search. "BLM" are estimates based on the 2-periods model.