A Distributional Framework for Matched Employer Employee Data
(Preliminary)

Interactions - BFI

Bonhomme, Lamadon, Manresa

University of Chicago MIT Sloan

September 26th - 2015

Wage Dispersion

- Wages are an important part of household income in the U.S., especially for households in the lower part of the income distribution.
- Understanding the sources of wage dispersion is a long standing question in Economics with important implications for inequality.
- Typically, only 30% of the variation in wages is explained through observable characteristics (e.g. Mortensen, 2001).

Matched Employer Employee Data

- Matched Employer Employee dataset follow workers over time and firms, keeping track both of worker and firm identifiers.
- Often merged with firm characteristics such as size, industry, etc. and worker characteristics such as education, age, gender, etc.
- These type of dataset open the possibility to account for unobserved worker and firm heterogeneity in addition to observables.

Fixed Effects Regression

A widely used methodology

- Abowd Kramarz Margolis (1999) (AKM) pioneered the use of matched data using a two-way fixed effect framework:

$$log(wage) = workerFE + firmFE + covar + error.$$

- AKM first documented the separate contribution of worker unobservables ($workerFE$), firm unobservables ($firmFe$), and the correlation between the two, $cov(workerFE, firmFE)$, to the overall wage variation.

- The AKM methodology is popular in many other settings involving sorting (e.g. Finkelstein, Gentzkow, Williams, 2014).

Fixed Effects Regression

Some drawbacks

Theory perspective:

- Imposes restrictive forms of complementary between firm and worker unobservables.
- At odds with large theoretical literature on sorting models with frictions.
  - Becker (1974), Shimer and Smith (2000), Eeckhout and Kircher (2011), Hagedorn Law Manovskii (2014), etc.

Empirics:

- Suffers from significant finite sample biases in panels with limited job mobility (Andrews et al. 2008, 2012).

This paper

- We propose a framework that allows for flexible forms of complementary between worker and firm unobservables, as well as sorting of workers into firms.
- We adopt a discrete heterogeneity approach: 'grouped fixed effects' for firms, and 'random effects' for workers.
- We recover wage distributions conditional on firm and worker types, as well as worker composition per firm.
- We show non-parametric identification when $T = 2$ or 4, and propose a 3-step estimation method.
- The framework is consistent with a number of theoretical sorting models.
- Key source of variation: wages of job movers (as in AKM).

Outline

1. Framework
2. Identification and Estimation
3. Application to Swedish Matched Employer Employee Data

Economic Environment and Wages in Period 1

- A worker i has a discrete type $\omega(i) \in \{1, \ldots, K\}$.
- A firm j belongs to a discrete class $f_j \in \{1, \ldots, L\}$.
- The indicator j_{it} is the firm where i works at time t.
- A worker of type k working in a firm of class ℓ draws from a wage distribution $F_{k\ell}$.
- The proportion of workers of type k in firm of class ℓ is $\pi_k(\ell)$.
- In this talk we abstract from worker and firm covariates.

Job Mobility: Two Periods Model

Period 1

- A worker of type k in a firm of class ℓ receives a wage Y_{i1} drawn from the distribution $F_{k\ell}$.

Period 2

- The worker moves to a firm of class ℓ' with a probability that depends on k, ℓ, and ℓ', but not on Y_{i1}.
- If the worker moves, then she draws a wage Y_{i2} from a distribution $G_{k\ell'}$ that depends on k, ℓ', but not on (ℓ, Y_{i1}).

Job Mobility: Four Periods Model

Periods 1 and 2

- A worker of type k draws wages (Y_{i1}, Y_{i2}) in a firm class ℓ from a bi-variate distribution that only depends on k and ℓ.

Period 3

- The worker moves to a firm of class ℓ' with a certain probability that depends on k, ℓ, Y_{i2}, but not Y_{i1}.
- If she moves, she receives a wage Y_{i3} that depends on k, ℓ, ℓ', Y_{i2} but not Y_{i1}.

Period 4

- The worker draws a wage Y_{i4} from a distribution that depends on k, ℓ', Y_{i3} but not on ℓ, Y_{i2}, Y_{i1}.

Comments

Assumptions

- The two periods model has two underlying assumptions:
  1. Job mobility is only driven by the type of the worker, k, and the classes of the firms, ℓ and ℓ' (but not on the draw itself).
  2. Serial independence upon job change.
- Both of these assumptions are relaxed in the four periods model, where
  1. Mobility depends, not only on k, ℓ and ℓ' but also on match-specific wage in period 2, Y_{i2}.
  2. Serial correlation upon job change.

Comments

Link Theoretical Models

Two Periods Model

- Consistent with theoretical models with no specific role for match-specific draws, unless independent over time or measurement error.
- Examples: Shimer and Smith (2000), with or without on-the-job search (worker threat's points being the value of unemployment).

Four Periods Model

- Models where state variables (k, f_t, Y_t) are 1st order Markov.
- Examples: Contract posting, sequential auctions (Lamadon, Lise, Meghir and Robin, 2015), with aggregate shocks (Lise and Robin, 2014).
- No latent human capital accumulation (kt), no permanent+transitory within-job wage dynamics.

Statistical Implications

- Two periods model: let $m_i = \mathbf{1}\{j_{i1} \neq j_{i2}\}$ denote job mobility:

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

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

- Four periods model: let $m_i = \mathbf{1}\{j_{i2} \neq j_{i3}\}$ denote job mobility:

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

  Type proportions of job movers are $p_k(y_{i2}, y_{i3}, \ell, \ell')$.

Empirical Equations - Two period model

In the two periods model we have that:

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

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

Empirical Equations - Two period model

In the two periods model we have that:

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

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

- Firm classes $f(j)$ can be recovered using (MOV) and/or (CROSS).
- $F_{k\ell}$, $G_{k\ell'}$ and $p_k(\ell, \ell')$ can be recovered from (MOV) using job movers. After, $\pi_k(\ell)$ can be recovered from (CROSS) using the cross-section.

Derivation

Empirical Equations - Four periods model

- In the more general four periods model we have that:

$$\Pr\left[Y_{i1} \leq y, Y_{i4} \leq y' | Y_{i2} = y_2, Y_{i3} = y_3, f(j_{i2}) = \ell, f(j_{i3}) = \ell', m_i = 1\right] = \sum_{k=1}^{K} p_k\left(y_2, y_3, \ell, \ell'\right) G_{k\ell}(y|y_2) H_{k\ell'}(y'|y_3).$$

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

Identification and Estimation Overview

We focus the discussion on the two periods model.

- Identification: three steps:
  1. Identify f_j the firm class of each firm.
  2. Identify $F_{k\ell}$ and $p_k(\ell, \ell')$ from joint distribution of movers conditional on firm classes.
  3. Identify $\pi_{k\ell}$ from cross section using firm classes and $F_{k\ell}$.

- Estimation: will follow the steps of identification.

Identification Wage functions

Simple Model (I)

Consider the following model for wages with matched effects:

$$Y_{i1} = a(\ell) + b(\ell)\alpha_i + \sigma(\ell)\varepsilon_{i1}.$$

Hence, a worker that changes jobs from firm of class ℓ to ℓ' draws wages according to:

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

Even after normalizing $a(\ell) = 0$ and $b(\ell) = 1$ (w.l.g.), this model is not nonparametrically identified using mean and covariance restrictions (e.g. Reiersol 1950).

Identification Wage functions

Simple Model (II)

Consider a worker that does the opposite transition: from ℓ' to ℓ:

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

We compare the difference in mean wages both in ℓ and ℓ' of workers doing opposite transitions:

$$\begin{aligned} E\left[Y_{i1} \,|\, \ell, \ell'\right] - E\left[Y_{i2} \,|\, \ell', \ell\right] &= b(\ell)\left(E\left[\alpha_i \,|\, \ell, \ell'\right] - E\left[\alpha_i \,|\, \ell', \ell\right]\right) \\ E\left[Y_{i2} \,|\, \ell, \ell'\right] - E\left[Y_{i1} \,|\, \ell', \ell\right] &= b(\ell')\left(E\left[\alpha_i \,|\, \ell, \ell'\right] - E\left[\alpha_i \,|\, \ell', \ell\right]\right) \end{aligned}$$

We identify the relative importance of the firm effects on wages for each firm provided that $E\left[\alpha_i \,|\, \ell, \ell'\right] - E\left[\alpha_i \,|\, \ell', \ell\right] \neq 0$ as:

$$\frac{E\left[Y_{i1} \,|\, \ell, \ell'\right] - E\left[Y_{i2} \,|\, \ell', \ell\right]}{E\left[Y_{i2} \,|\, \ell, \ell'\right] - E\left[Y_{i1} \,|\, \ell', \ell\right]} = \frac{b(\ell)}{b(\ell')}.$$

Identification Wage Functions

Simple Model (III)

Empirical counterpart for $E\left[\alpha_i \,|\, \ell, \ell'\right] - E\left[\alpha_i \,|\, \ell', \ell\right] \neq 0$:

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

Normalizing $a(\ell) = 0$ and $b(\ell) = 1$, $a(\ell')$ can be recovered solving:

$$\begin{aligned} E\left[Y_{i1} \,|\, \ell, \ell'\right] &= E\left[\alpha_i \,|\, \ell, \ell'\right] \\ E\left[Y_{i2} \,|\, \ell, \ell'\right] &= a(\ell') + b(\ell')E\left[\alpha_i \,|\, \ell, \ell'\right]. \end{aligned}$$

For any other ℓ', $a(\ell'')$ and $b(\ell'')$ can be recovered using movements (back and forth) from ℓ' to either ℓ or ℓ'.

Identification Wage function

More General Case

- Consider the stationary case. Let $\ell \neq \ell'$ be firm classes, and write (MOV) in **matrix form** for (ℓ, ℓ') and (ℓ', ℓ) transitions:

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

  where $A(\ell, \ell')$ has as elements $\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, \ell, \ell'\right]$.

- The identification argument relies on a **joint diagonalization** of these matrices.

- It is sufficient (but not necessary, see paper) for identification of $F_{k\ell}$ that:
  - $p_k(\ell', \ell) \neq 0$ for $k = 1, ..., K$.
  - $\frac{p_k(\ell, \ell')}{p_k(\ell', \ell)}$, $k = 1, ..., K$, are distinct.
  - The columns of $F(\ell)$ (the $F_{k\ell}$) are linearly independent.

Firm Classes

Classification Problem

- 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.

- When adding information from **job movers**, conditions needed for identification are weaker.

Identification of Worker Composition

- Given firm classes $f(j)$ and cdfs $F_{k\ell}$ we can recover **type proportions** $\pi_k(\ell)$ from cross-sectional wages.
- Writing
$$H(\ell) = F(\ell)\Pi(\ell),$$
where $H(\ell)$ has elements $\Pr\left[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell\right]$, the $K \times 1$ vector $\Pi(\ell)$ has elements $\pi_k(\ell)$.
- Since the columns of $F(\ell)$ have rank K we can obtain $\Pi(\ell)$ by inversion.
- Transition probabilities are also identified:
$$\Pr\left[f(j_{i2}) = \ell' \,|\, \omega(i) = k, f(j_{i1}) = \ell, mi = 1\right] = \frac{p_k\left(\ell, \ell'\right) q_{\ell,\ell'}}{\sum_{\ell'} p_k\left(\ell, \ell'\right) q_{\ell,\ell'}}.$$

3 Step Estimation Method

1. Use firm - specific wage distributions to estimate class firms.
2. Given estimates of firm classes, use joint distribution of wages in t = 1 and t = 2 from job movers to recover $F_{k\ell}$ using a conditional mixture model.
3. Given estimates of classes and $F_{k\ell}$ use cross-sectional distribution to obtain $\pi_k(\ell)$.

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 (2014) even when the number of firms grows substantially faster than the number of workers per firm.

2nd step: Wage functions

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

- Under parametric assumptions on $F_{k\ell}$, we maximize the following Likelihood with respect to θ:

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

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

3rd step: Worker Composition

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

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

- This log-likelihood function is **concave**. In practice we use a second EM algorithm.

- Alternatively, $\pi_k(\ell)$ could be estimated by solving a constrained linear regression problem.

Sample description (I)

- 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) and in 2004 (period 2). We refer to this sample of 800,000 workers and 50,000 firms as Sample 1.

Sample description (II)

- From Sample 1 we select job movers who are full-year employed and whose firm IDs are different in 2002 and 2004.
- In order to avoid considering job changes that are unrelated to job mobility we discard workers whose firm identifier is not present in 2002 or 2004.
- Sample 2 is the resulting sample of 20,000 **job movers**, with 13,000 firms. We also estimated the model with 55,000 movers without imposing the above restriction.
- In both samples we compute log pre-tax **annual earnings** net of time dummies (within education*cohort).

Estimated Firm Classes

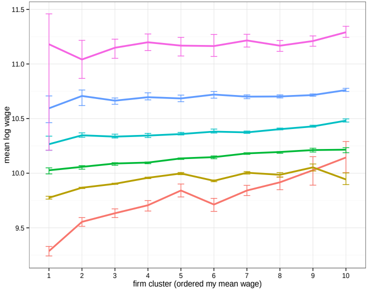
firm cluster:	1	2	3	4	5	6	7	8	9	10	all
number of workers	21,662	62,929	110,792	114,324	100,080	78,837	137,971	85,806	58,728	27,023	798,152
number of firms	6,487	7,972	7,804	6,494	4,663	3,748	4,209	3,984	3,157	2,812	51,330
% HS dropout	28.9	28	26.6	26.9	23.7	21.1	18.9	12.2	5.31	3	20.7
% HS grade	59.7	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: Estimated firm classes on the 2002 wave of Sample 1, using a weighted k-means algorithm (empirical cdfs with 40 points, 1000 starting values). All workers are males, employed during the full year 2002. "HS" is high school.

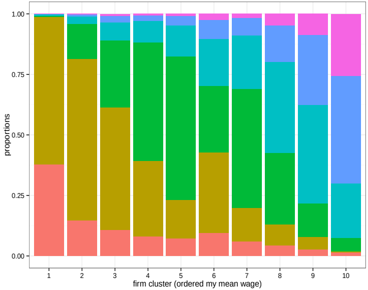
Estimated Mean log Wages and Proportions

2 Period Model

Mean log wages



Proportions of worker types



Notes: The left 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 colors. 95% confidence intervals based on the parametric bootstrap (200 replications). The right graph plots type proportions* $\widehat{\pi}_k(\ell)$.

Simulation and Decomposition

- 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 (Two 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.

Estimation of the four periods model

- We add two employment periods: 2001 and 2005. There are 12,519 workers moving in 2003 employed in the five years.
- We estimate the following model for movers between periods 2 and 3:

$$
\begin{aligned}
Y_{i1} &= \rho_{12}Y_{i2} + \mu_{1,k\ell} + \sigma_{1,k\ell}\varepsilon_{i1}, \\
Y_{i2} &= \mu_{2,k\ell} + \xi_{\ell'} + \sigma_{2,k\ell}\varepsilon_{i2}, \\
Y_{i3} &= \rho_{32}Y_{i2} + \mu_{3,k\ell'} + \xi_{\ell} + \sigma_{3,k\ell'}\varepsilon_{i3}, \\
Y_{i4} &= \rho_{43}Y_{i3} + \mu_{4,k\ell'} + \sigma_{4,k\ell'}\varepsilon_{i4},
\end{aligned}
$$

with ε's standard normal and independent. We first estimate the parameters for job movers, then for stayers, using the EM algorithm.

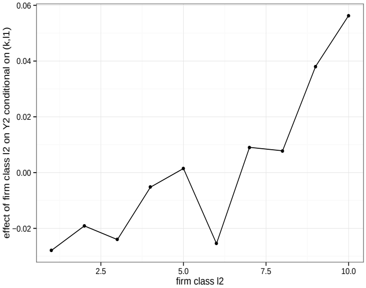
Variance decompositions on Swedish data (four periods model, preliminary)

constraint	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	ρ_{32}	ρ_{12}	lik. (movers)
	Data						
linear	0.2841	0.5904	0.1254	0.1532	0.65	0.861	25650.17
monotonic	0.2629	0.6482	0.0888	0.1076	0.674	0.87	26087.19
none	0.2587	0.6599	0.0814	0.0985	0.693	0.876	26307.9
exogenous	0.7766	0.0473	0.1762	0.4598	0	0	-
	Simulated from the monotonic model						
exogenous	0.8019	0.0353	0.1628	0.4834	0	0	-

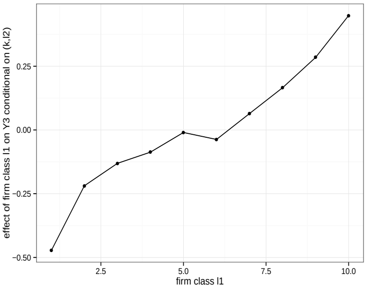
Notes: α is the worker type fixed-effect, ψ is the firm class fixed-effect. "lik movers" is the value of the log-likelihood in the sample of job movers.

Effects of firm classes on wages (four periods model, preliminary)

Effect of ℓ' on Y_{i2} (given k, ℓ)



Effect of ℓ on Y_{i3} (given k, ℓ')



Notes: Average effects, four periods model with monotonicity in k. The effect of mobility m on Y_{i2}, given (k, ℓ), is 0.042 (not shown on the left graph).

Conclusions

- Econometric framework for matched data
  - Useful to study wage variation in a context with complementarities between unobservables.
  - Other interesting applications of the method: sorting across schools, neighborhoods, cities, hospitals, etc.
- 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.
- Currently working on on statistical properties when heterogeneity may not be grouped in the population, and clustering provides an approximation to the structure of heterogeneity.

Derivation

$$
\begin{aligned}
&\Pr\left[Y_{i1} \le y\,,\, Y_{i2} \le y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] = \\
&= \sum_{k=1}^{K} \Pr\left[Y_{i1} \le y\,,\, Y_{i2} \le y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1, \omega(i) = k\right] p_k\left(\ell, \ell'\right) \\
&= \sum_{k=1}^{K} \Pr\left[Y_{i1} \le y \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1, \omega(i) = k\right] \cdot \\
&\qquad \cdot \Pr\left[Y_{i2} \le y \,|\, Y_{i1} \le y, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1, \omega(i) = k\right] \cdot \\
&\qquad\quad \cdot p_k\left(\ell, \ell'\right) \\
&= \sum_{k=1}^{K} \Pr\left[Y_{i1} \le y \,|\, f(j_{i1}) = \ell, \omega(i) = k\right] \cdot \\
&\qquad \cdot \Pr\left[Y_{i2} \le y \,|\, f(j_{i2}) = \ell', m_i = 1, \omega(i) = k\right] \\
&\qquad\quad \cdot p_k\left(\ell, \ell'\right) \\
&= \sum_{k=1}^{K} p_k\left(\ell, \ell'\right) F_{k\ell}(y) F_{k\ell'}(y').
\end{aligned}
$$

go back

Identification Wage function

More General Case

- 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.

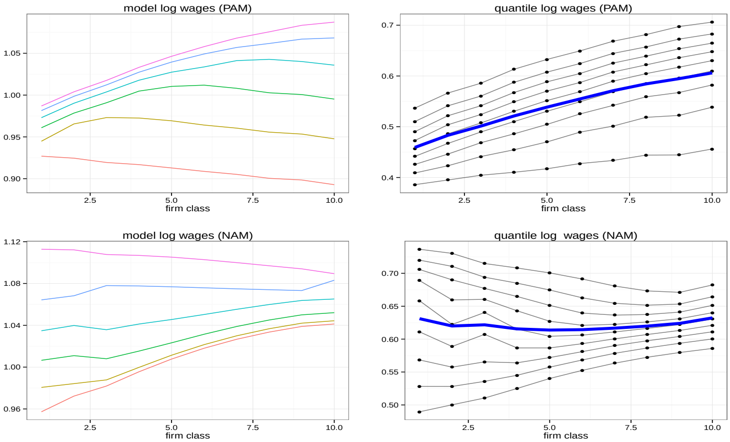
- Letting $W_1 = S^{-\frac{1}{2}}U'$ and $W_2 = S^{-\frac{1}{2}}V'$, we have:

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

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.

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 (Two periods model)

		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 two periods model.