The Anatomy of the Wage Distribution. How do Gender and Immigration Matter?

Rasmus Lentz* Suphanit Piyapromdee† Jean-Marc Robin‡

April 15, 2016

Preliminary and incomplete

Abstract

Workers with similar observed characteristics may have different wage paths because their unobserved characteristics are rewarded differently or because they have different mobility patterns. For example, controlling for observed characteristics, a native worker may have a steeper wage profile than his counterpart immigrant because he has different unobserved characteristics or because he is more likely to be employed at a more productive firm (likewise for male versus female workers). Understanding the contributions of worker and firm heterogeneity to wage and mobility differentials can shed light on many key economic issues such as wage inequality, statistical discrimination and wage assimilation of immigrants.

In this paper, we propose an estimation method that allows for unrestricted interactions between worker and firm unobserved characteristics in both wages and the moving probability. Related to Bonhomme, Lamadon and Manresa (2014) (BLM), our method identifies double sided unobserved heterogeneity through an application of the EM-algorithm where the firm classification is repeatedly updated so as to improve on the likelihood function. In Monte Carlo simulations we demonstrate that the cyclic updating of the firm classification provides a significant performance improvement.

Keywords: On-the-job search; Heterogeneity; Wage distributions

JEL codes: E24; E32; J63; J64

*University of Wisconsin-Madison. E-mail: rlentz@wisc.edu

†University College London. Email s.piyapromdee@ucl.ac.uk

‡Sciences-Po, Paris and University College London. Email jmarc.robin@gmail.com

1 Introduction

The studies of inequality and discrimination are both vexed by the challenges of measuring the returns to unobservable characteristics and the correlation between observables and unobservables. The challenge is further complicated by the fact that the relevant unobservables reside on both sides of the employment relationship; with both firms and workers. For example, are women paid less because they are women? Or because their gender is correlated with particular worker side unobservables? Or because they tend to work for firms or in jobs that pay less regardless of the gender of the worker?

In this paper, we propose an estimation method that allows for unrestricted interactions between worker and firm unobserved characteristics in both wages and the mobility patterns. Related to Bonhomme et al. (2015) (BLM), our method identifies double sided unobserved heterogeneity through an application of the EM-algorithm.

The log wage of a match is assumed to be normally distributed with a mean and variance that depends on both observed and unobserved worker and firm heterogeneity. Unobserved heterogeneity is described through groupings where all workers in a given group are identical according to the unobserved characteristics, similarly for firms.

Workers move between firms according to transition probabilities that depend both on worker and firm characteristics. Employed workers meet other employment opportunities at a rate that depends on the type of the current firm. The type of the employment opportunity is drawn from a distribution that is common to all workers and independent of the type of the current firm. Finally, the probability that the worker decides to quit the current job and move to the new firm is assumed to depend both on the worker's type as well as the types of the two firms involved. In our particular specification, the choice probability takes the shape of a binomial logit. Hence, one interpretation of the mobility patterns in the paper is that of a standard on the job search model with random utility.

The estimator maximizes the likelihood of observing workers' wage and employment histories given a data structure that records the identity of the worker's employer at any point in time as well non-employment. We do not consider the earliest part of a worker's history and furthermore assume the worker's state at the beginning of the observation window is drawn from the steady state. Wages are observed for each employment spell and can change within the spell at an annual frequency. Unobserved heterogeneity on the worker side is modeled as a random effect whereas the firm side is modeled as a fixed effect because the likelihood function evaluation is infeasible in case of a firm side random effect. The maximization of the likelihood of the data is implemented through an application of the EM-algorithm with double sided unobserved heterogeneity.

The estimator is initialized by some firm classification. We choose to follow BLM and group firms by the k-means algorithm using wage data. The estimator proceeds to update the firm classification in the iteration steps. For a given firm classification, the algorithm performs an EM step where a worker classification (in terms of a posterior on a worker's type) is obtained in the E step for given wage and mobility parameters, and in the M-step the wage and mobility parameters are updated subject to the worker and firm classifications so as to maximize worker posterior conditional expected log-likelihood of the data. The mobility parameters are close to those of a Bradley-Terry model but nevertheless involve a substantial non-linearity in the likelihood function with respect to the mobility parameters. We formulate an MM algorithm that allow a fast solution for mobility parameters that improve on the likelihood function - a substantial contribution for the feasibility of the estimator. With the likelihood improved by the EM step, the firm classification is then occasionally updated before a new EM step so as to improve on the expected log-likelihood. The firm classification updates are done subject to the entire likelihood of the data and can change substantially from its initialization.

The model will be estimated on Danish matched employer-employee data where the worker's employment match state is observed at a weekly frequency. The wage data are observed specifically for the given match in question and can change at an annual frequency. We observe both hours and earnings in the match. The model allows us to provide a detailed measurement of the returns to observable characteristics such as gender, ethnic origin, age, and education. And given the dynamic structure of the model we can also distinguish between wage and value variance and their decompositions.

2 The Model

Workers are indexed by $i \in \{1,...,I\}$ and firms are indexed by $j \in \{0,1,...,J\}$, where $j = 0$ reflects unemployment. For each worker i in week t, we observe (w_{it}, j_{it}, x_{it}) where $t = 1,2,\dots,T_i$ and $j_{it} \equiv j(i,t) \in \{0,1,...,J\}$ is the ID of the firm employing worker i at time t, x_{it} are observed worker controls, and w_{it} is the workers wage rate at time t. Wages are measured at annual frequency within a given match. Thus, wage observations are missing except for the first week of the match and the first week of the year. The presentation of the estimator suppresses the dependence on the controls. In the case where the model parameters are fully flexible in the controls, we implement the controls by stratification of the data.

We assume that firms can be clustered into L different groups indexed by $\ell \in \{1,\dots,L\}$ and that workers can be clustered into K different groups indexed by $k \in \{1,...,K\}$. The index ℓ_j is the type of firm j and k_i is the type of worker i. We shall be treating the unobserved firm types $\mathscr{L} = (\ell_1,...,\ell_J)$ as a a fixed effect (i.e. a parameter to be estimated) and the worker type as a random effect mixing conditional distributions of workers' trajectories. Unemployment is observable and is denoted by $\ell = 0$. Let $\pi = (\pi_1, ..., \pi_K)$ denote the proportions of workers' types in the population.

Let $f_{k\ell}(w)$ denote the wage density, conditional on the worker's type k and the employer's type ℓ. In the expressions below, adopt the convention that $f_{k0}(\cdot) = 1$ since there is no wage attached to unemployment spells. We assume w_{it} and $w_{it'}$ independent conditional on $\ell_{j(i,t)}, \ell_{j(i,t')}$. We further denote the probability for a worker of type k of making a transition from a firm of type ℓ to a firm of type ℓ' (possibly $\ell' = \ell$) at time t as $M_{k\ell\ell'}$, and the probability of staying with the same employer is $\overline{M}_{k\ell} = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}$. With this, $M_{k\ell 0}$ is the layoff rate in a match between a type k worker and type ℓ firm. The state of unemployment is special in that by definition, $M_{k00} = 0$. The worker type k conditional job finding rate in the model is $1 - \overline{M}_{k0}$. The wage density can easily accommodate time varying controls. In the case of the transition probabilities we only allow fixed control since the estimation uses stationarity. The stationarity assumption allows us to make inference about mobility parameters not only from worker flows, but also from the initial allocation of workers to jobs.

3 EM algorithm

In describing the estimation algorithm, we suppress the dependence on controls.

3.1 Likelihood with observed firm types and stationarity

Let us first consider the case where the employers' types ℓ_j are observed. Let $\ell_{it} = \ell_{j(i,t)}$ denote, by some abuse of notation, the type of the firm employing worker i in period t. Let also

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

indicate an employer change between t and $t+1$.

For a value $\beta = (f, M, \pi)$ of the parameters and a classification $\mathscr{L}$ of firms, the likelihood for one worker i is

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

where $L_i(k; \beta, \mathscr{L})$ is the individual likelihood conditional on worker type k, i.e.

$$L_i(k; \beta, \mathscr{L}) = \pi_k m_{k,\ell_{i1}} \prod_{t=1}^{T_i - 1} f_{k,\ell_{it}}(w_{it}) \prod_{t=1}^{T_i - 1} \overline{M}_{k,\ell_{it}}^{1-D_{it}} M_{k,\ell_{it},\ell_{i,t+1}}^{D_{it}}, \tag{1}$$

where by construction $f(w_{it}) = 1$ if the wage observation is missing. All the spell likelihoods are multiplied to form the likelihood of observing worker i's spell history.

Note that $m_{k,\ell(i,1)}$ is the probability that a type k worker is with a type-ℓ_{i1} firm at the first observation period $t = 1$. These probabilities cannot be assumed structurally independent of the transition probabilities as the identification of sorting relies on mobility patterns. We thus link $m_{k\ell}$ to transition probabilities $M_{k\ell\ell'}$ by assuming them to be equal to the stationary probabilities solving the linear system:

$$m_{k\ell} = m_{k\ell}\overline{M}_{k\ell} + \sum_{\ell'=0}^{L} m_{k\ell'} M_{k,\ell',\ell}. \tag{2}$$

In matrix notations, $\mathbf{m}_k = [m_{k\ell}]_\ell$ is the eigenvector associated to the eigenvalue one of the transition matrix $\mathbf{M}_k^\top$ where:

$$\mathbf{M}_k = \begin{pmatrix} \overline{M}_{k0} & M_{k01} & \cdots & M_{k0L} \\ M_{k10} & \overline{M}_{k1} + M_{k11} & \cdots & M_{k1L} \\ \vdots & \vdots & \ddots & \vdots \\ M_{kL0} & M_{kL1} & \cdots & \overline{M}_{kL} + M_{kLL} \end{pmatrix}.$$

Alternatively, using the constraint $\mathbf{1}_L^\top \mathbf{m}_k = 1$, where $\mathbf{1}_L = \text{ones}(L+1,1)$ is the $L+1$-vector of ones, one can rewrite $\mathbf{m}_k$ as the solution to

$$\mathbf{1}_L = \mathbf{B}_k \mathbf{m}_k, \quad \mathbf{B}_k = (\mathbf{J}_L + \mathbf{M}_k^\top) - \mathbf{I}_L,$$

where $\mathbf{J}_L = \text{ones}(L+1, L+1)$. Hence, $\mathbf{m}_k = \mathbf{B}_k^{-1}\mathbf{1}_L$. This procedure guaranties that $\mathbf{m}_k$ sums to one and is a distribution of probability masses.

3.2 A cyclic EM algorithm

The EM algorithm (Dempster et al., 1977) consists of iterating the calculation of the posterior probabilities of worker types (E-step) and the expected log-likelihood maximization using the type probabilities calculated in the E-step (M-step). See Arcidiacono and Jones (2003); Bonhomme and Robin (2009); Arcidiacono and Miller (2011) for recent applications in economics. The firm classification is in the data unobserved. It is infeasible to evaluate the likelihood function for the formulation of the model where a firm's unobserved type is a latent variable symmetric to the unobservable worker type formulation in equation (1). The difficulty lies with accounting for the codependency between a firm's workers resulting from their matches to a common firm type in a setup where workers move between firms. Consequently, we adopt a fixed effect approach to the firm classification. The estimate of the firm classification $\mathscr{L}$ is updated so as to improve on the likelihood of the data as stated in equation (1).

For a given value of $\beta = (f, M, \pi)$, the posterior probability of worker i to be of type k given all wages and controls (all the available information) is

$$p_i(k;\beta,\mathscr{L}) = \frac{L_i(k;\beta,\mathscr{L})}{\sum_{k=1}^{K} L_i(k;\beta,\mathscr{L})}. \tag{3}$$

Then, define

$$Q_i(f,\mathscr{L};\beta^{(m)},\mathscr{L}^{(m)}) = \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \left[\sum_{t=1}^{T_i} \ln f_{k,\ell(i,t)}(w_{it})\right]. \tag{4}$$

Q_i is the worker posterior conditional expected log-likelihood given worker i's wage and employer history for a given value $\widetilde{\beta}$ of the parameter. The worker posteriors are determined by the model parameters and firm classification $\left(\beta^{(m)},\mathscr{L}^{(m)}\right)$, where the superscript is used to denote the given EM-algorithm iteration. Also, let

$$\begin{aligned} H_i(M,\mathscr{L};\beta^{(m)},\mathscr{L}^{(m)}) = \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \Bigg[& \ln m_{k,\ell(i,1)} \\ & + \sum_{t=1}^{T_i-1} \left\{ (1-D_{it}) \ln \overline{M}_{k,\ell(i,t)} + D_{it} \ln M_{k,\ell(i,t),\ell(i,t+1)} \right\} \Bigg]. \end{aligned} \tag{5}$$

H_i is the expected log-likelihood given worker i's wage and employer history conditional on the worker posteriors. Finally, The EM algorithm iterates the following steps.

E-step For $\beta^{(m)} = (f^{(m)}, M^{(m)}, \pi^{(m)})$ and $\mathscr{L}^{(m)}$ calculate posterior probabilities $p_i(k;\beta^{(m)},\mathscr{L}^{(m)})$.

M-step Update $\beta^{(m)}$ by maximizing $\sum_i \sum_k p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \ln L_i(k;\beta,\mathscr{L}^{(m)})$ subject to $\sum_k \pi_k = 1$, that is

$$f^{(m+1)} = \arg\max_f \sum_{i=1}^{I} Q_i(f,\mathscr{L}^{(m)};\beta^{(m)},\mathscr{L}^{(m)}), \tag{6}$$

$$M^{(m+1)} = \arg\max_M \sum_{i=1}^{I} H_i(M,\mathscr{L}^{(m)};\beta^{(m)},\mathscr{L}^{(m)}), \tag{7}$$

$$\pi_k^{(m+1)} = \frac{1}{I} \sum_{i=1}^{I} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}). \tag{8}$$

C-step Update $\mathscr{L}^{(m)}$ by maximizing $\sum_i \sum_k p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \ln L_i(k;\beta^{(m+1)},\mathscr{L})$ with respect to

$\mathscr{L}$, that is

$$\mathscr{L}^{(m+1)} = \arg \max_{\mathscr{L}=(\ell_1,...,\ell_J)} \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \left[\sum_{t=1}^{T_i} \ln f_{k,\ell_{j(i,t)}}^{(m+1)}(w_{it}) \right]$$
$$+ \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \left[\ln m_{k,\ell_{j(i,1)}}^{(m+1)} + \sum_{t=1}^{T_i-1} \left\{ (1-D_{it}) \ln \overline{M}_{k,\ell_{j(i,t)}}^{(m+1)} + D_{it} \ln M_{k,\ell_{j(i,t)},\ell_{j(i,t+1)}}^{(m+1)} \right\} \right]. \tag{9}$$

This is a cyclic EM algorithm because in the M-step we maximize with respect to β given $\mathscr{L}^{(m)}$ and in the C-step we maximize with respect to $\mathscr{L}$ given $\beta^{(m+1)}$. In practice we shall iterate the E and M steps keeping the firm classification constant until convergence (or for a few iterations) before running the C-step. The C-step is computationally too costly to be repeated too often.

3.3 Firm classification

In order to initialize the algorithm we follow Bonhomme et al. (2015) and cluster firms with similar observed characteristics. Let z_j denote a vector of characteristics for firm j. For example, z_j may contain the deciles of the distribution of all wages w_{it} such that $j(i,t) = j$, the average job duration at firm j, etc. One can then use any standard classification algorithm, such as k-means for example, to classify firms into L groups.

Next, the C-step is a very difficult optimization problem because of the high dimensionality of the control variable $\mathscr{L} = (\ell_1,...,\ell_J)$. Complexity stems from the fact that the posterior likelihood in (9) is non separable in $\mathscr{L} = (\ell_1,...,\ell_J)$ because of the terms $M_{k,\ell_{j(i,t)},\ell_{j(i,t+1)}}$. We suggest to iterate the following firm-specific optimizations,

$$\ell_j^{(p+1)} = \arg\max_{\ell_j} \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)})) \left[\sum_{t=1}^{T_i} \ln f_{k,\ell_j}^{(m+1)}(w_{it}) \times \mathbf{1}\{j(i,t)=j\} \right]$$
$$+ \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \left[\ln m_{k,\ell_j}^{(m+1)} \times \mathbf{1}\{j(i,1)=j\} \right]$$
$$+ \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)}) \left[\sum_{t=1}^{T_i-1} \left\{ (1-D_{it}) \ln \overline{M}_{k,\ell_j}^{(m+1)} + D_{it} \ln M_{k,\ell_j,\ell_{j(i,t+1)}^{(p)}}^{(m+1)} \right\} \times \mathbf{1}\{j(i,t)=j\} \right]. \tag{10}$$

In practice it suffices to iterate until we find a value $\mathscr{L}^{(m+1)}$ such that

$$\sum_{i=1}^{I}\sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)})\ln L_i(k;\beta^{(m+1)},\mathscr{L}^{(m+1)}) > \sum_{i=1}^{I}\sum_{k=1}^{K} p_i(k;\beta^{(m)},\mathscr{L}^{(m)})\ln L_i(k;\beta^{(m+1)},\mathscr{L}^{(m)}).$$

The observed state ℓ_0, unemployment, is trivially assigned group $\ell = 0$.

4 Empirical specification

4.1 Wage distribution

Wages are assumed lognormal given match type. Specifically,

$$f_{k\ell}(w) = \frac{1}{\sigma_{k\ell}}\varphi\left(\frac{w-\mu_{k\ell}}{\sigma_{k\ell}}\right), \tag{11}$$

with $\varphi(x) = (2\pi)^{-1/2}e^{-x^2/2}$. This specification of the log-wage mean allows for a match-specific mean $\mu_{k\ell}$ and variance $\sigma^2_{k\ell}$.

The M-step update 6 takes the following form:

$$\mu_{k\ell}^{(m+1)} = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}w_{it}}{\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}}$$

$$\sigma_{k\ell}^{(m+1)} = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}[w_{it}-\mu_{k\ell}^{(m+1)}]^2}{\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}}.$$

4.2 Transition probabilities

The probability for a worker of type k of a transition from a firm of type $\ell \geq 0$ to a firm of type $\ell' \geq 0$ at time t is

$$M_{k\ell\ell'} = \lambda_\ell \nu_{\ell'} P_{k\ell\ell'}. \tag{12}$$

Parameter $\lambda_\ell \in [0,1]$ is the probability of a meeting with an outside employer when the current state is either unemployment or a job of type ℓ. Parameter $\nu_{\ell'} \geq 0$, with $\sum_{\ell'=0}^{L}\nu_{\ell'} = 1$, is the probability that the outside draw is unemployment or a job of type ℓ'.

The parameter $P_{k\ell\ell'}$ is the probability that the transition from ℓ to ℓ' becomes effective. We assume a Bradley-Terry specification for $P_{k\ell\ell'}$ (see e.g. Agresti, 2003; Hunter, 2004). That is,

$$P_{k\ell\ell'} = \frac{\gamma_{k\ell'}}{\gamma_{k\ell}+\gamma_{k\ell'}}\mathbf{1}\{\ell \neq 0 \vee \ell' \neq 0\}. \tag{13}$$

Parameter $\gamma_{k\ell}$, with $\sum_{\ell=0}^{L}\gamma_{k\ell}=1$, measures the quality of the match (k,ℓ). If the worker draws a same-type job we assume the workers moves with probability 1/2.[1] For a currently unemployed worker such same-type transitions do not make sense. A draw $\ell'=0$ when $\ell=0$ never generates an observed mobility. So $P_{k00}=0$.

We assume that the offer distribution out of non-employment is the same as that when employed. Since offers are not always accepted, we do not trivially obtain the offer distribution from the distribution of accepted jobs out of non-employment. Rather the assumption is the basis for the identification of γ_{k0} and λ_0. Note yet that a worker can change job within the same sector ℓ; it makes no sense for unemployment.

With this it follows that, for $\ell \geq 0$,

$$\overline{M}_{k\ell} = 1-\sum_{\ell'=0}^{L} M_{k\ell\ell'} = 1-\lambda_\ell \sum_{\ell'=0}^{L} \nu_{\ell'} P_{k\ell\ell'} = 1-\lambda_\ell+\lambda_\ell \sum_{\ell'=0}^{L} \nu_{\ell'}(1-P_{k\ell\ell'}).$$

4.3 M-step update for transition probabilities

In the M-step of the EM algorithm, we maximize the part of the expected likelihood that refers to transitions, i.e.

$$\widetilde{H}(M;\beta^{(m)}) \equiv \sum_{k=1}^{K}\sum_{\ell=0}^{L}\left\{\overline{n}_{k\ell}(\beta^{(m)})\ln\overline{M}_{k\ell}+\sum_{\ell'=0}^{L} n_{k\ell\ell'}(\beta^{(m)})\ln M_{k\ell\ell'}\right\},$$

where

$$\overline{n}_{k\ell}(\beta^{(m)}) = \sum_i p_i(k;\beta^{(m)})\,\#\{t: D_{it}=0, \ell_{it}=\ell\},$$
$$n_{k\ell\ell'}(\beta^{(m)}) = \sum_i p_i(k;\beta^{(m)})\,\#\{t: D_{it}=1, \ell_{it}=\ell, \ell_{i,t+1}=\ell'\},$$

where $\#\{\}$ denotes the cardinality of a set. This likelihood is similar to the likelihood of a Bradley-Terry model except that when the incumbent firm ℓ wins we do not know against which ℓ'. The likelihood is thus rendered more nonlinear by the presence of the term in $\ln\overline{M}_{k\ell}$. An MM algorithm can still be developed as follows.

Because the logarithm is concave, we can minorize $\ln\overline{M}_{k\ell}$ as follows. With obvious notations,

[1] We experimented the specification

$$M_{k,\ell,\ell'} = \lambda_\ell \nu_{\ell'} \frac{\gamma_{k,\ell'}}{\theta\gamma_{k\ell}+\gamma_{k,\ell'}},$$

where $\theta>0$ measures the incumbent's advantage and parametrizes mobility within the same group of firms. However it appeared difficult to disentangle θ from λ.

for $\ell \geq 0$,

$$\begin{aligned}\ln \overline{M}_{k\ell} &= \ln\left(1-\lambda_\ell + \sum_{\ell'=0}^{L} \lambda_\ell \nu_{\ell'}(1-P_{k\ell\ell'})\right)\\ &\geq \frac{1-\lambda_\ell^{(s)}}{\overline{M}_{k\ell}^{(s)}} \ln\left(\frac{1-\lambda_\ell}{1-\lambda_\ell^{(s)}}\overline{M}_{k\ell}^{(s)}\right) + \sum_{\ell'=0}^{L} \frac{\lambda_\ell^{(s)}\nu_{\ell'}^{(s)}(1-P_{k\ell\ell'}^{(s)})}{\overline{M}_{k\ell}^{(s)}} \ln\left(\frac{\lambda_\ell \nu_{\ell'}(1-P_{k\ell\ell'})}{\lambda_\ell^{(s)}\nu_{\ell'}^{(s)}(1-P_{k\ell\ell'}^{(s)})}\overline{M}_{k\ell}^{(s)}\right).\end{aligned}$$

Note that both sides of the inequality are equal if $\beta = \beta^{(s)}$ (no parameter change).

Let

$$\widetilde{n}_{k\ell\ell'}^{(s)} = \overline{n}_{k\ell}^{(m)} \frac{\lambda_\ell^{(s)}\nu_{\ell'}^{(s)}(1-P_{k\ell\ell'}^{(s)})}{\overline{M}_{k\ell}^{(s)}}.$$

This is the predicted number of times that home beats visitor ℓ'. Given initial values $\lambda_\ell^{(s)}, \nu_{\ell'}^{(s)}$ one can update $\gamma^{(s)}$ so as to maximize

$$\sum_{k=1}^{K}\sum_{\ell=0}^{L}\sum_{\ell'=0}^{L}\left\{\widetilde{n}_{k\ell\ell'}^{(s)} \ln \frac{\gamma_{k\ell}}{\gamma_{k\ell}+\gamma_{k\ell'}} + n_{k\ell\ell'} \ln \frac{\gamma_{k\ell'}}{\gamma_{k\ell}+\gamma_{k\ell'}}\right\},$$

subject to the normalization $\sum_{\ell=0}^{L} \gamma_{k\ell} = 1$. Now, because

$$-\ln(\gamma_{k\ell}+\gamma_{k\ell'}) \geq 1 - \ln(\gamma_{k\ell}^{(s)}+\gamma_{k\ell'}^{(s)}) - \frac{\gamma_{k\ell}+\gamma_{k\ell'}}{\gamma_{k\ell}^{(s)}+\gamma_{k\ell'}^{(s)}}$$

(see Hunter, 2004), we can instead maximize

$$\sum_{k=1}^{K}\sum_{\ell=0}^{L}\left(\sum_{\ell'=0}^{L}(\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell'\ell})\right)\ln\gamma_{k\ell} - \sum_{k=1}^{K}\sum_{\ell=0}^{L}\sum_{\ell'=0}^{L}\left((\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'})\frac{\gamma_{k\ell}+\gamma_{k\ell'}}{\gamma_{k\ell}^{(s)}+\gamma_{k\ell'}^{(s)}}\right).$$

That is (taking special care to indices),

$$\gamma_{k\ell}^{(s+1)} \propto \left(\sum_{\ell'=0}^{L}(\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell'\ell})\right)\left[\sum_{\ell'=0}^{L}\frac{\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'} + \widetilde{n}_{k\ell'\ell}^{(s)} + n_{k\ell'\ell}}{\gamma_{k\ell}^{(s)}+\gamma_{k\ell'}^{(s)}}\right]^{-1},$$

where $X_\ell \propto Y_k$ means $X_\ell = Y_\ell / \sum_\ell Y_\ell$, that is $\gamma_{k\ell}^{(s+1)}$ should sum to one over $\ell \geq 0$.

Update $\lambda^{(s)}$ by maximizing

$$\sum_{\ell=0}^{L}\left(\sum_{k=1}^{K}\left(\overline{n}_{k\ell}\frac{1-\lambda_\ell^{(s)}}{\overline{M}_{k\ell}^{(s)}}\right)\ln(1-\lambda_\ell) + \sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'}\right)\ln\lambda_\ell\right).$$

Let

$$\lambda_{\ell}^{(m+1)} = \left[\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(m)} + n_{k\ell\ell'}^{(m)}\right)\right]\left[\sum_{k=1}^{K}\left(\overline{n}_{k\ell}^{(m)}\frac{1-\lambda_{\ell}^{(m)}}{\overline{M}_{k\ell}^{(m)}}\right) + \sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(m)} + n_{k\ell\ell'}^{(m)}\right)\right]^{-1}.$$

Finally update $\nu^{(s)}$ by maximizing

$$\sum_{\ell'=0}^{L}\left[\sum_{k=1}^{K}\sum_{\ell=0}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'}\right)\right]\ln \nu_{\ell'} \quad \text{s.t.} \quad \sum_{\ell'=0}^{L}\nu_{\ell'} = 1.$$

That is

$$\nu_{\ell}^{(s+1)} = \frac{\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left[\widetilde{n}_{k\ell'\ell}^{(s,m)} + n_{k\ell'\ell}^{(m)}\right]}{\sum_{\ell=0}^{L}\sum_{k=1}^{K}\sum_{\ell'=0}^{L}\left[\widetilde{n}_{k\ell'\ell}^{(s,m)} + n_{k\ell'\ell}^{(m)}\right]}.$$

5 Monte Carlo simulations

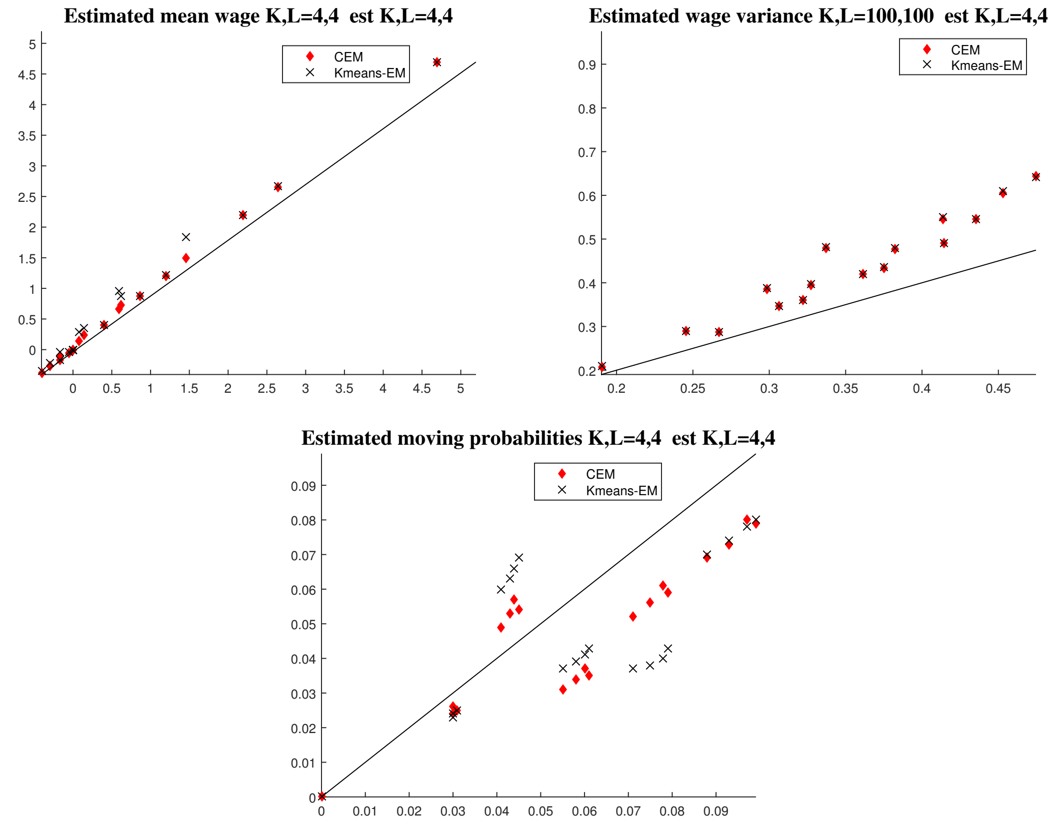
We illustrate the performance of the estimator by its ability to recover the parameters used to generate a simulated data set. To simplify, all heterogeneity in the simulation is unobserved. A match with a type x worker and type y firm produces surplus $S(x,y) = \exp\left[xy - \alpha(x-y)^2\right]$. Assume the log wage to be proportional to surplus according to, $w(x,y) = \ln[S(x,y)/2] + \varepsilon$, where ε is normally distributed with variance σ_{xy}^2. The probability of preferring a firm y to a firm z is $S(x,y)/[S(x,y)+S(x,z)]$.

There are K worker types, x_k, $k \in \{1,\ldots,K\}$ and L firm types, y_ℓ, $\ell \in \{1,\ldots,L\}$. Let the distribution of types be uniform over both worker and firm types. Approximate x and y to come from a log-normal distribution. Hence, set $\ln x_k = \Phi^{-1}(0.2+0.6k/K)$ and $\ln y_\ell = \Phi^{-1}(0.2+0.6\ell/L)$, where Φ is the standard normal CDF and c is a positive constant to ensure that x_K and y_L are finite numbers. Furthermore, specify $\lambda_\ell = \lambda^0 + \lambda^1\ell/L$ and $\nu_\ell \propto \nu^0 + \nu^1\ell/L$. Finally, specify the type dependent wage variance by, $\sigma_{x_k y_\ell}^2 = \omega_0 + \omega_x k/K + \omega_y \ell/L$. In the simulations we work with an economy that has 50,000 workers and 5,000 firms simulated over a 5 year period.

5.1 Matching to finite types

We adopt the following parameterization. Let the true number of firm and worker types, $K = L = 4$. The surplus function is parameterized by, $\alpha = 0.1$. Furthermore, $(\lambda_0,\lambda_1) = (0.5, 0.2)$ and $(\nu_0,\nu_1) = (0.13, 0.2)$. Finally, $(\omega_0,\omega_x,\omega_y) = (0.1, 0.2, 0.2)$. The estimation of the model is done subject to the true number of worker and firm types. This is done partly to facilitate comparison of the model estimates with the true parameters. But it also avoids the complicated issue of how

Figure 1: Comparison of True and Estimated Parameters



to choose the number of unknown types, which is particularly challenging in regards to the firm classification. We return to this issue in the next section.

To illustrate how the firm type reclassification in the c-step impacts our results, we report two sets of estimates that we summarize in Figure 1 as well as Tables 3-6. The top left panel in Figure 1 shows the estimated mean wage parameters relative to the true values. The horizontal axis shows the true value and the vertical axis holds the estimated value. The red diamonds show the estimates from the cyclic EM algorithm, whereas the gray x's show the estimates from an EM algorithm subject to the initial k-means firm classification. The k-means clustering results in 24 percent of misclassified firm types. After about 10 iterations, this proportion of misclassified firm types reduces to 10 percent along with an improvement in the likelihood function.

While both estimation algorithms perform well, the cyclic EM performs substantially better. We can choose a simple equally weighted average of absolute differences as a distance metric. With such a metric, the distance between the CEM estimates of $\mu_{k\ell}$ relative to the true parameters is 0.032. The distance between the k-means based EM $\mu_{k\ell}$ estimates and the truth is 0.109, a factor of 3.4 greater than the CEM estimates. The top right hand panel shows the CEM and k-means EM estimates of $\sigma^2_{k\ell}$ relative to the truth in the same fashion. The distance between the

CEM estimates of $\sigma^2_{k\ell}$ and the truth is 0.019, whereas it is 0.054 in the case of the k-means based EM algorithm. The lower panel shows the performance of the estimator in terms of the estimated mobility probabilities $M_{k\ell\ell'}$. The CEM algorithm again performs better. The distance between CEM estimates and the truth is 0.033. The k-means based EM estimates have a distance from the truth of 0.046. The tables provide the actual values of the parameter estimates.

5.1.1 Monte Carlo Repetitions

So as to eliminate simulation noise from the analysis we repeat the Monte Carlo exercise in Section 5.1 S times where the simulated data is redone with a different random number generator seed for each simulation. The results are done for $S = 100$. Denote by $\mu^s_{k\ell}$ the mean wage estimate for simulation s. Use the same superscript notation for the other model parameter estimates. We show both the CEM algorithm estimator as well as the k-means EM estimator.

Table 1: CEM parameter estimates

CEM $E[\mu^s_{k\ell}]$				CEM Var$[\mu^s_{k\ell}]$				k-means EM $E[\mu^s_{k\ell}]$			
-0.376	-0.248	-0.140	-0.090	0.000	0.002	0.002	0.005	-0.351	-0.158	-0.054	-0.197
-0.092	0.191	0.506	0.731	0.001	0.001	0.004	0.068	-0.031	0.426	0.851	0.335
0.292	0.788	1.420	1.909	0.003	0.050	0.169	0.284	0.383	1.246	2.147	1.056
0.877	1.766	3.105	4.113	0.018	0.225	0.715	1.199	0.980	2.734	4.598	2.119

CEM $E\left[\left(\sigma^2_{k\ell}\right)^s\right]$				CEM Var$\left[\left(\sigma^2_{k\ell}\right)^s\right]$				k-means EM $E\left[\left(\sigma^2_{k\ell}\right)^s\right]$			
0.216	0.268	0.312	0.334	0.000	0.000	0.000	0.001	0.229	0.304	0.349	0.291
0.278	0.321	0.361	0.389	0.000	0.000	0.000	0.001	0.298	0.357	0.400	0.359
0.352	0.382	0.413	0.458	0.000	0.000	0.000	0.001	0.383	0.414	0.449	0.451
0.470	0.455	0.486	0.587	0.001	0.000	0.001	0.032	0.519	0.508	0.504	0.677

Table 1 presents the average wage parameter CEM estimates and the variance of the CEM estimates over the S simulation repetitions. It furthermore shows the k-means EM wage parameter estimates. The true model parameters can be found in Table 4. While both estimators perform well, the CEM estimator performs substantially better than k-means EM estimator. Taking an equally weighted absolute difference over all the model parameter estimates, the CEM estimator has an average distance from the true model parameters over the 100 simulation repetitions of 6.18. The same distance measure has value 16.81 in the case of the k-means EM estimator.

Table 2: Match value parameter estimates

True $\gamma_{k\ell}$					CEM $E[\gamma^s_{k\ell}]$				
0.158	0.175	0.196	0.222	0.250	0.150	0.176	0.196	0.233	0.244
0.106	0.130	0.166	0.230	0.367	0.100	0.136	0.174	0.259	0.332
0.051	0.072	0.112	0.207	0.558	0.047	0.077	0.136	0.276	0.464
0.007	0.014	0.033	0.108	0.837	0.007	0.017	0.057	0.265	0.654

k-means $E[\gamma^s_{k\ell}]$				
0.146	0.217	0.159	0.224	0.254
0.095	0.163	0.172	0.323	0.248
0.045	0.090	0.182	0.488	0.195
0.007	0.017	0.145	0.756	0.074

In Table 2 we show the average of the $\gamma_{k\ell}$ estimates for the CEM and k-means EM and compare it to the true values. It is apparent that the CEM provides a significant improvement over the initial k-means classification.

5.2 The number of worker and firm types

We now briefly discuss the difficulty of estimating the number of unkown types in the estimation. The fixed effect approach to the firm classification complicates the application of Information Criterion approaches such as Akaike or Bayesian. The comparison of the likelihood across different number of firm types turns out to be of little value. To see this, consider a simple example where there is only a single firm and worker type and there is no unemployment. Hence, in truth, the steady state probability is fully placed on a single combination $\widetilde{m}_{k\ell_{it}} = 1$. But the estimation divides firms arbitrarily into $L = 2$, half of the firms in one group and half in the other. The model parameters are the same across the two firm types since there is no difference between them. However, in the evaluation of the likelihood in equation (1) it follows that $m_{k\ell_{it}} = 1/2$. The likelihood of the worker's mobility pattern is similarly reduced. Thus, the likelihood mechanically drops as a result of the arbitrary split of a firm group into two identical groups. Of course, had the likelihood understood that a firm is equally likely to be one or the other firm type, the likelihood would have remained unchanged. But the fixed effect approach to the firm classification does not allow this consideration.

For L known, it is in principle possible to determine the number of worker types, K, by an Information Criterion approach, properly formulated for our data. But the optimal K depends on the given choice of L which is set somewhat arbitrarily. Hence, we will instead adopt a robustness approach to the number of types in our estimation with a focus on the sensitivity of the estimated wage and mobility parameters.

5.3 Aggregation and misclassification

In the case where the true worker and type space is continuous, the estimation will necessarily approximate the type heterogeneity in the data with a courser grouping. Even when the true type space is finite, the same may happen. In order to facilitate a comparison between the true model parameters and the estimate, we calculate averages of the true model for the aggregation implied by the estimation.

Let k,ℓ refer to the true groups and let $\widehat{k},\widehat{\ell}$ index estimated ones. The estimation produces the worker i posterior, $p_i(\widehat{k})$. Denote by, $\widehat{\pi}_{\widehat{k}} = \frac{1}{I}\sum_i p_i(\widehat{k})$ the estimated proportion of type-$\widehat{k}$ workers. Let $\pi_k = \frac{\#\{k_i=k\}}{I}$ be the true proportion of type-k workers. In addition, define

$$\widehat{p}(\widehat{k}|k) = \frac{1}{I}\sum_{i:k_i=k} p_i(\widehat{k})/\pi_k$$

as the estimated proportion of type-$\widehat{k}$ workers given true type k. Lastly, let

$$p(k|\widehat{k}) = \frac{\widehat{p}(\widehat{k}|k)\pi_k}{\widehat{\pi}_{\widehat{k}}}$$

be the probability that the worker is of true type k given estimated type $\widehat{k}$.

In the same way for the firm classification, define

$$\widehat{q}(\widehat{\ell}|\ell) = \frac{\#\{\ell_j=\ell,\widehat{\ell}_j=\widehat{\ell}\}}{\#\{\ell_j=\ell\}}, \quad q(\ell|\widehat{\ell}) = \frac{\#\{\ell_j=\ell,\widehat{\ell}_j=\widehat{\ell}\}}{\#\{\widehat{\ell}_j=\widehat{\ell}\}}$$

as the estimated probability that a true type ℓ firm is of type $\widehat{\ell}$, and the probability that an estimated type $\widehat{\ell}$ is of true type ℓ, respectively.

With this, define the following aggregations. For the wage function parameters, define the group average over the true parameters,

$$\widetilde{\mu}_{\widehat{k}\widehat{\ell}} = \sum_k p(k|\widehat{k})\sum_\ell q(\ell|\widehat{\ell})\mu_{k\ell}, \quad \widetilde{\sigma}^2_{\widehat{k}\widehat{\ell}} = \sum_k p(k|\widehat{k})\sum_\ell q(\ell|\widehat{\ell})\sigma^2_{k\ell}. \tag{14}$$

Assuming independence, $p(k|\widehat{k})q(\ell|\widehat{\ell})$ is the probability that a match between a worker that is estimated to be type $\hat{k}$ and a firm that is estimated to be type $\hat{\ell}$ is in fact a match between a true type k worker and a true type ℓ firm, in which case the wage in a match would be governed by wage parameters $\left(\mu_{k\ell},\sigma^2_{k\ell}\right)$. In the following we compare the model estimates $\hat{\mu}_{\widehat{k}\widehat{\ell}}$ with $\widetilde{\mu}_{\widehat{k}\widehat{\ell}}$ where the latter is considered the proper comparison given the classification error embodied in the estimation.

The same comparison is done for the estimated transition probabilities,

$$\widetilde{M}_{\widehat{k}\widehat{\ell}\widehat{\ell}'} = \sum_k p(k|\widehat{k}) \sum_\ell q(\ell|\widehat{\ell}) \sum_{\ell'} M_{k\ell\ell'} \widehat{q}(\widehat{\ell}'|\ell'),$$

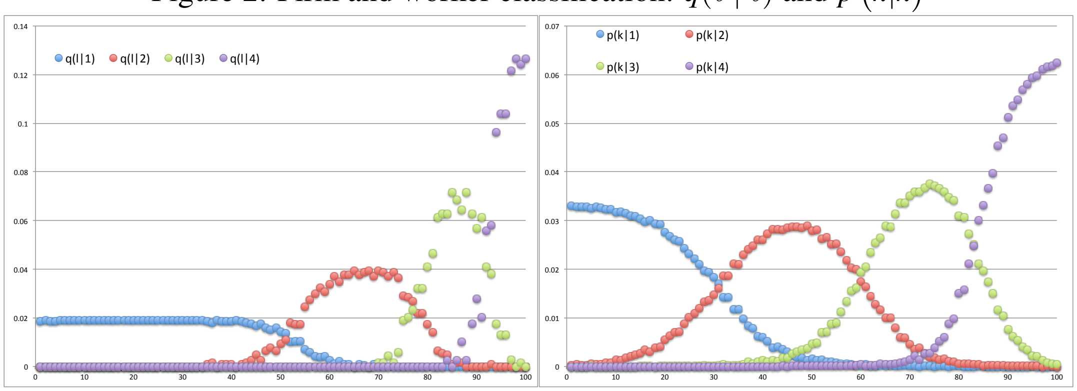
where $\widetilde{M}_{\widehat{k}\widehat{\ell}\widehat{\ell}'}$ represents the true model parameters given the classification error.

5.4 Matching to a "continuous" type space.

In this example, we increase the number of true firm and worker types to $K = L = 100$. The surplus function is parameterized by, $\alpha = 0.1$. Furthermore, $(\lambda_0, \lambda_1) = (0.5, 0.2)$ and $(\nu_0, \nu_1) = (0.13, 0.2)$. Finally, $(\omega_0, \omega_x, \omega_y) = (0.1, 0.2, 0.2)$. Table (7) reports the estimate of wage parameters $\mu^w_{k\ell}, \sigma^w_{k\ell}$ when $\hat{K} = \hat{L} = 4$ in the right panel, and the aggregated true parameters calculated using (14) in the left panel.

Overall, the estimates are close to the aggregated true parameters,. Tables 7 and 9 show the estimated and aggregated true values for the wage parameters and transition probabilities. As can be seen there is a very tight relationship between the true parameters and the estimated parameters for the given misclassification of worker and firm types in the estimation. But of course a major part of the estimation is the identification of the worker and firm classifications. Figure 2 illustrates

Figure 2: Firm and worker classification: $q(\ell \mid \hat{\ell})$ and $p\left(k|\hat{k}\right)$



the classification error in the estimation through the functions, $q(\ell|\widehat{\ell})$ and $p(k|\widehat{k})$. As can be seen, an estimated type is broadly speaking associated with a connected subset of types in the true type space. This is true for both firms and workers. In this example, the classification is performing very well.

One can as before calculate the distance between the estimated model parameters and the true model parameters, where the latter is represented by $\widetilde{\mu}_{\widehat{k}\widehat{\ell}}$ in the case of the mean wage parameters. For the wage parameters, μ and σ^2, the CEM and k-means EM algorithms performance are roughly equivalent. For the μ estimates, the CEM distance is 0.086 and the k-means EM distance is 0.088.

For the σ^2 estimates, the CEM distance is 0.082 and the k-means EM distance is 0.083. For the mobility parameter estimates, the CEM algorithm performs substantially better. Here, the CEM distance is 0.026 and the k-means EM distance is 0.05.

6 Estimation

TBC

7 Discrimination and Inequality

TBC

8 Conclusion

TBC

Figures and Tables

Table 3: Comparison of True and Estimated Wage Parameters using Kmeans EM

true $\mu^w_{k\ell}$				estimated $\mu^w_{k\ell}$			
-0.406	-0.295	-0.167	-0.050	-0.355	-0.218	-0.166	-0.047
-0.166	0.078	0.400	0.869	-0.040	0.286	0.407	0.871
0.146	0.587	1.198	2.189	0.358	0.950	1.224	2.190
0.616	1.453	2.645	4.690	0.877	1.834	2.670	4.687

true $\sigma^w_{k\ell}$				estimated $\sigma^w_{k\ell}$			
0.200	0.250	0.300	0.350	0.225	0.287	0.305	0.348
0.250	0.300	0.350	0.400	0.292	0.355	0.353	0.403
0.300	0.350	0.400	0.450	0.382	0.469	0.408	0.451
0.350	0.400	0.450	0.500	0.508	0.681	0.490	0.498

Table 4: Comparison of True and Estimated Wage Parameters using Cyclic EM

true $\mu^w_{k\ell}$				estimated $\mu^w_{k\ell}$			
-0.406	-0.295	-0.167	-0.050	-0.381	-0.272	-0.168	-0.048
-0.166	0.078	0.400	0.869	-0.108	0.133	0.401	0.869
0.146	0.587	1.198	2.189	0.245	0.664	1.209	2.189
0.616	1.453	2.645	4.690	0.725	1.487	2.661	4.687

true $\sigma^w_{k\ell}$				estimated $\sigma^w_{k\ell}$			
0.200	0.250	0.300	0.350	0.212	0.260	0.304	0.347
0.250	0.300	0.350	0.400	0.269	0.312	0.350	0.402
0.300	0.350	0.400	0.450	0.350	0.393	0.405	0.451
0.350	0.400	0.450	0.500	0.412	0.453	0.476	0.498

Table 5: Comparison of True and Estimated Transition Probabilities using Kmeans EM

	true $M_{k\ell\ell'}, k=1$					estimated $M_{k\ell\ell'}, k=1$				
	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$
$\ell=0$	0.000	0.041	0.055	0.071	0.088	0.000	0.060	0.037	0.037	0.070
$\ell=1$	0.030	0.043	0.058	0.075	0.093	0.023	0.063	0.039	0.038	0.074
$\ell=2$	0.030	0.044	0.060	0.078	0.097	0.024	0.066	0.041	0.040	0.078
$\ell=3$	0.031	0.045	0.061	0.079	0.099	0.025	0.069	0.043	0.043	0.080
$\ell=4$	0.031	0.045	0.061	0.080	0.100	0.025	0.068	0.042	0.041	0.080

	true $M_{k\ell\ell'}, k=2$					estimated $M_{k\ell\ell'}, k=2$				
$\ell=0$	0.000	0.043	0.061	0.083	0.111	0.000	0.064	0.043	0.045	0.091
$\ell=1$	0.028	0.043	0.062	0.085	0.116	0.021	0.063	0.043	0.045	0.095
$\ell=2$	0.026	0.041	0.060	0.085	0.118	0.019	0.058	0.041	0.041	0.094
$\ell=3$	0.023	0.037	0.055	0.079	0.115	0.020	0.060	0.042	0.043	0.093
$\ell=4$	0.018	0.029	0.044	0.066	0.100	0.014	0.046	0.034	0.033	0.080

	$M_{k\ell\ell'}, k=3$					estimated $M_{k\ell\ell'}, k=3$				
$\ell=0$	0.000	0.046	0.069	0.098	0.132	0.000	0.060	0.037	0.037	0.070
$\ell=1$	0.026	0.043	0.067	0.099	0.140	0.023	0.063	0.039	0.038	0.074
$\ell=2$	0.021	0.037	0.060	0.095	0.143	0.024	0.066	0.041	0.040	0.078
$\ell=3$	0.014	0.026	0.046	0.079	0.136	0.025	0.069	0.043	0.043	0.080
$\ell=4$	0.007	0.013	0.023	0.046	0.100	0.025	0.068	0.042	0.041	0.080

	$M_{k\ell\ell'}, k=4$					estimated $M_{k\ell\ell'}, k=4$				
$\ell=0$	0.000	0.052	0.082	0.114	0.142	0.000	0.064	0.053	0.067	0.119
$\ell=1$	0.021	0.043	0.077	0.118	0.155	0.021	0.063	0.058	0.076	0.137
$\ell=2$	0.012	0.028	0.060	0.112	0.166	0.009	0.028	0.041	0.065	0.149
$\ell=3$	0.005	0.012	0.030	0.079	0.165	0.004	0.014	0.024	0.043	0.123
$\ell=4$	0.001	0.002	0.005	0.020	0.100	0.001	0.002	0.006	0.012	0.080

Table 6: Comparison of True and Estimated Transition Probabilities using Cyclic EM

	true $M_{k\ell\ell'}, k=1$					estimated $M_{k\ell\ell'}, k=1$				
	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$
$\ell=0$	0.000	0.037	0.045	0.052	0.059	0.000	0.035	0.038	0.039	0.029
$\ell=1$	0.033	0.043	0.052	0.060	0.069	0.028	0.042	0.047	0.048	0.036
$\ell=2$	0.037	0.049	0.060	0.069	0.080	0.032	0.050	0.055	0.057	0.043
$\ell=3$	0.042	0.056	0.069	0.079	0.091	0.036	0.055	0.062	0.064	0.049
$\ell=4$	0.046	0.062	0.075	0.087	0.100	0.040	0.062	0.070	0.074	0.058

	true $M_{k\ell\ell'}, k=2$					estimated $M_{k\ell\ell'}, k=2$				
$\ell=0$	0.000	0.036	0.040	0.041	0.036	0.000	0.038	0.038	0.032	0.023
$\ell=1$	0.034	0.043	0.049	0.051	0.046	0.026	0.042	0.043	0.036	0.026
$\ell=2$	0.041	0.052	0.060	0.063	0.058	0.032	0.053	0.055	0.048	0.035
$\ell=3$	0.049	0.063	0.074	0.079	0.075	0.040	0.065	0.070	0.064	0.048
$\ell=4$	0.059	0.078	0.093	0.102	0.100	0.044	0.072	0.079	0.075	0.058

	$M_{k\ell\ell'}, k=3$					estimated $M_{k\ell\ell'}, k=3$				
$\ell=0$	0.000	0.033	0.034	0.029	0.017	0.000	0.032	0.029	0.022	0.013
$\ell=1$	0.036	0.043	0.045	0.039	0.024	0.030	0.042	0.040	0.032	0.019
$\ell=2$	0.045	0.055	0.060	0.055	0.035	0.038	0.056	0.055	0.046	0.029
$\ell=3$	0.056	0.072	0.081	0.079	0.056	0.045	0.069	0.072	0.064	0.043
$\ell=4$	0.070	0.093	0.112	0.120	0.100	0.050	0.079	0.085	0.082	0.058

	$M_{k\ell\ell'}, k=4$					estimated $M_{k\ell\ell'}, k=4$				
$\ell=0$	0.000	0.030	0.026	0.016	0.004	0.000	0.026	0.019	0.010	0.004
$\ell=1$	0.038	0.043	0.039	0.026	0.007	0.034	0.042	0.035	0.021	0.008
$\ell=2$	0.050	0.060	0.060	0.044	0.014	0.044	0.060	0.055	0.036	0.016
$\ell=3$	0.064	0.082	0.091	0.079	0.032	0.052	0.078	0.081	0.064	0.033
$\ell=4$	0.077	0.104	0.129	0.141	0.100	0.056	0.089	0.099	0.093	0.058

Table 7: Comparison of Aggregated True and Estimated Wage Parameters using Kmeans EM

aggregated true $\mu^w_{k\ell}$				estimated $\mu^w_{k\ell}$			
-0.265	0.029	0.229	0.354	-0.291	-0.010	0.191	0.336
-0.024	0.501	0.897	1.177	0.018	0.507	0.919	1.275
0.274	1.121	1.790	2.286	0.391	1.161	1.831	2.406
0.564	1.782	2.765	3.509	0.829	1.948	2.925	3.723

aggregated true $\sigma^w_{k\ell}$				estimated $\sigma^w_{k\ell}$			
0.190	0.267	0.306	0.328	0.209	0.288	0.347	0.396
0.245	0.322	0.362	0.383	0.290	0.361	0.421	0.478
0.298	0.375	0.415	0.436	0.389	0.435	0.492	0.546
0.337	0.414	0.453	0.474	0.481	0.549	0.609	0.642

Table 8: Comparison of Aggregated True and Estimated Wage Parameters using CEM

aggregated true $\mu^w_{k\ell}$				estimated $\mu^w_{k\ell}$			
-0.265	0.029	0.229	0.354	-0.292	-0.011	0.189	0.334
-0.024	0.501	0.897	1.177	0.015	0.504	0.914	1.271
0.274	1.121	1.790	2.286	0.386	1.159	1.825	2.402
0.564	1.782	2.765	3.509	0.829	1.947	2.916	3.722

aggregated true $\sigma^w_{k\ell}$				estimated $\sigma^w_{k\ell}$			
0.190	0.267	0.306	0.328	0.208	0.287	0.347	0.395
0.245	0.322	0.362	0.383	0.289	0.361	0.420	0.478
0.298	0.375	0.415	0.436	0.386	0.434	0.491	0.546
0.337	0.414	0.453	0.474	0.479	0.547	0.606	0.643

Table 9: Comparison of Aggregated True and Estimated Transition Probabilities using Kmeans

	aggregated true $M_{k\ell\ell'}, k=1$					estimated $M_{k\ell\ell'}, k=1$				
	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$
$\ell=0$	0.000	0.115	0.088	0.059	0.038	0.000	0.104	0.075	0.048	0.031
$\ell=1$	0.001	0.118	0.091	0.061	0.039	0.001	0.093	0.071	0.048	0.031
$\ell=2$	0.001	0.115	0.090	0.062	0.040	0.001	0.093	0.072	0.049	0.032
$\ell=3$	0.001	0.109	0.086	0.060	0.039	0.001	0.089	0.069	0.048	0.031
$\ell=4$	0.001	0.104	0.083	0.058	0.038	0.001	0.085	0.066	0.046	0.030

	aggregated true $M_{k\ell\ell'}, k=2$					estimated $M_{k\ell\ell'}, k=2$				
$\ell=0$	0.000	0.122	0.099	0.068	0.044	0.000	0.105	0.079	0.052	0.034
$\ell=1$	0.001	0.118	0.099	0.070	0.046	0.001	0.093	0.078	0.055	0.037
$\ell=2$	0.001	0.101	0.090	0.067	0.046	0.001	0.081	0.072	0.053	0.036
$\ell=3$	0.001	0.084	0.078	0.060	0.042	0.001	0.068	0.062	0.048	0.033
$\ell=4$	0.001	0.071	0.068	0.053	0.038	0.001	0.059	0.055	0.043	0.030

	aggregated true $M_{k\ell\ell'}, k=3$					estimated $M_{k\ell\ell'}, k=3$				
$\ell=0$	0.000	0.131	0.112	0.077	0.050	0.000	0.103	0.083	0.054	0.036
$\ell=1$	0.001	0.119	0.110	0.080	0.053	0.001	0.093	0.089	0.065	0.044
$\ell=2$	0.001	0.084	0.091	0.074	0.052	0.001	0.063	0.072	0.059	0.042
$\ell=3$	0.000	0.056	0.066	0.060	0.045	0.000	0.040	0.052	0.048	0.036
$\ell=4$	0.000	0.039	0.049	0.047	0.038	0.000	0.029	0.039	0.038	0.030

	aggregated true $M_{k\ell\ell'}, k=4$					estimated $M_{k\ell\ell'}, k=4$				
$\ell=0$	0.000	0.140	0.123	0.083	0.053	0.000	0.100	0.086	0.057	0.037
$\ell=1$	0.001	0.120	0.120	0.087	0.057	0.001	0.093	0.101	0.072	0.048
$\ell=2$	0.001	0.068	0.091	0.080	0.058	0.000	0.044	0.072	0.066	0.048
$\ell=3$	0.000	0.034	0.055	0.060	0.049	0.000	0.019	0.040	0.048	0.041
$\ell=4$	0.000	0.018	0.033	0.041	0.038	0.000	0.010	0.022	0.031	0.030

Table 10: Comparison of Aggregated True and Estimated Transition Probabilities using CEM

	aggregated true $M_{k\ell\ell'}, k=1$					estimated $M_{k\ell\ell'}, k=1$				
	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$	$\ell'=0$	$\ell'=1$	$\ell'=2$	$\ell'=3$	$\ell'=4$
$\ell=0$	0.000	0.115	0.088	0.059	0.038	0.000	0.104	0.075	0.048	0.031
$\ell=1$	0.001	0.118	0.091	0.061	0.039	0.001	0.093	0.071	0.048	0.031
$\ell=2$	0.001	0.115	0.090	0.062	0.040	0.001	0.092	0.072	0.049	0.032
$\ell=3$	0.001	0.109	0.086	0.060	0.039	0.001	0.089	0.070	0.048	0.031
$\ell=4$	0.001	0.104	0.083	0.058	0.038	0.001	0.084	0.066	0.046	0.030

	aggregated true $M_{k\ell\ell'}, k=2$					estimated $M_{k\ell\ell'}, k=2$				
$\ell=0$	0.000	0.122	0.099	0.068	0.044	0.000	0.105	0.079	0.052	0.033
$\ell=1$	0.001	0.118	0.099	0.070	0.046	0.001	0.093	0.078	0.055	0.037
$\ell=2$	0.001	0.101	0.090	0.067	0.046	0.001	0.081	0.072	0.053	0.036
$\ell=3$	0.001	0.084	0.078	0.060	0.042	0.001	0.068	0.063	0.048	0.033
$\ell=4$	0.001	0.071	0.068	0.053	0.038	0.001	0.059	0.055	0.043	0.030

	aggregated true $M_{k\ell\ell'}, k=3$					estimated $M_{k\ell\ell'}, k=3$				
$\ell=0$	0.000	0.131	0.112	0.077	0.050	0.000	0.103	0.083	0.055	0.035
$\ell=1$	0.001	0.119	0.110	0.080	0.053	0.001	0.093	0.089	0.065	0.044
$\ell=2$	0.001	0.084	0.091	0.074	0.052	0.001	0.063	0.072	0.059	0.042
$\ell=3$	0.000	0.056	0.066	0.060	0.045	0.000	0.041	0.052	0.048	0.036
$\ell=4$	0.000	0.039	0.049	0.047	0.038	0.000	0.029	0.039	0.038	0.030

	aggregated true $M_{k\ell\ell'}, k=4$					estimated $M_{k\ell\ell'}, k=4$				
$\ell=0$	0.000	0.140	0.123	0.083	0.053	0.000	0.100	0.086	0.057	0.037
$\ell=1$	0.001	0.120	0.120	0.087	0.057	0.001	0.093	0.100	0.072	0.048
$\ell=2$	0.001	0.068	0.091	0.080	0.058	0.000	0.044	0.072	0.066	0.048
$\ell=3$	0.000	0.034	0.055	0.060	0.049	0.000	0.019	0.040	0.048	0.041
$\ell=4$	0.000	0.018	0.033	0.041	0.038	0.000	0.010	0.022	0.031	0.030

A Simulation notes

Link structure and initialization

There are I individuals and J firms. First assign firm ID's to firm type classes. Assign a type $k \in \{1,...,K\}$ and a type $\ell \in \{1,...,L\}$ uniformly to all workers and firms. Then assign $\ln x = \Phi^{-1}(k/K)$ to all workers of type k and $\ln y = \Phi^{-1}(\ell/L)$ to all firms of class ℓ.

Let χ_ℓ be the vacancy intensity of firms type ℓ. In that case, the probability of drawing a type-ℓ firm is given by,

$$v_\ell = \frac{J_\ell \chi_\ell}{\sum_{\ell'} J_{\ell'} \chi_{\ell'}}.$$

For each worker draw an initial match firm type from $m_{k\ell}$, conditional on k. For the ℓ realization, draw a firm id j from group of ℓ type firm id's. With this procedure, there will be no simulation noise in the number of firms by firm class and there will be no noise in the number of workers by worker type class.

Durations, transitions and wages

Simulate spell lengths where the spell hazard in a job (k,ℓ) is given by $h_{k\ell} = \sum_{\ell'} M_{k\ell\ell'} = 1 - \overline{M}_{k\ell}$. First simulate the duration. The probability of a duration n is geometrically distributed, $\Pr(dur = T) = (1-h_{k\ell})^{T-1} h_{k\ell}$, with CDF $G(T) = 1-(1-h_{k\ell})^T$. Duration is found by inverting the CDF. Take a uniformly distributed random number, u, invert the

CDF,

$$u = 1-(1-h_{k\ell})^T \Leftrightarrow T \ln(1-h_{k\ell}) = \ln(1-u) \Leftrightarrow T = \frac{\ln(1-u)}{\ln(1-h_{k\ell})} + 1.$$

Conditional on the spell ending, then simulate the destination. The probability that the worker moves to a type ℓ' firm conditional on moving is $M_{k\ell\ell'}/h_{k\ell}$. After the fact, for each year of the duration of the simulated spell, simulate wages.

B Parametric specifications for the impact of observed heterogeneity.

As an alternative to the fully flexible specification in the main analysis where controls introduced by stratification of the data, instead consider a parametric specification where controls x_{it} are introduced into wages and mobility parameters as follows. As before, let wages be log-normal, but controls are imposed on the average by the parametric specification,

$$f_{k\ell}(w|\mathbf{x}) = \frac{1}{\sigma_{k\ell}}\varphi\left(\frac{w-\mu_{k\ell}(\mathbf{x})}{\sigma_{k\ell}}\right), \tag{15}$$

where,

$$\mu_{kl}(x \mid \mu) = \mu_{kl} + \mathbf{x}'\mu.$$

The M-step update of the μ parameters is done by weighted least squares,

$$\mu^{(m+1)} = \arg\min_{\mu} \frac{\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}\left[w_{it}-\mu_{kl}(x_{it} \mid \mu)\right]^2}{\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}}.$$

And the variance update is as before,

$$\sigma_{k\ell}^{(m+1)} = \frac{\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}[w_{it}-\mu_{kl}\left(x_{it} \mid \mu^{(m+1)}\right)]^2}{\sum_{i=1}^{I} p_i(k;\beta^{(m)})\sum_{t=1}^{T_i}\mathbf{1}\{j(i,t)=\ell\}}.$$

Denote by $\hat{x}_i$ a subset of time invariant controls. One approach to the estimation of mobility parameters would be make the mobility parameters be fully flexible in $\hat{x}$ and carry on as in the stratified case. In this case, the exclusion restrictions would only be imposed through the wage part of the estimation and in the choice of which controls to leave out of $\hat{x}$. This approach does not require any modification of the estimation method. For a simple set of controls such as gender, immigrant status, and education (say 3 education groups), this could be course enough to be an attractive approach.

References

Agresti, A. (2003). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley.

Arcidiacono, P. and J. B. Jones (2003, 05). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. *Econometrica 71*(3), 933–946.

Arcidiacono, P. and R. A. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica 79*(6), 1823–1867.

Bonhomme, S., T. Lamadon, and E. Manresa (2015). A Distributional Framework for Matched Employer Employee Data. mimeo, University of Chicago.

Bonhomme, S. and J.-M. Robin (2009, 01). Assessing the equalizing force of mobility using short panels: France, 1990-2000. *Review of Economic Studies 76*(1), 63–92.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(1), pp. 1–38.

Hunter, D. R. (2004, 02). Mm algorithms for generalized bradley-terry models. *Ann. Statist. 32*(1), 384–406.