On Worker and Firm Heterogeneity in Wages and Employment Mobility: Evidence from Danish Register Data

Rasmus Lentz* Suphanit Piyapromdee† Jean-Marc Robin‡

February 5, 2018

Preliminary and incomplete

Abstract

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. (2017) (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. Firm classification is a result of both wage and mobility patterns in the data. We estimate the model on Danish matched employer-employee data for the period 1985-2013. The estimation includes gender, education, age and time controls. We find an increased sorting pattern over time, although overall sorting is modest.

Keywords: Heterogeneity; Wage distributions; Employment and job mobility; Matched employer-employee data; Finite mixtures; EM algorithm; Classification algorithm; Sorting; Decomposition of wage inequality

JEL codes: E24; E32; J63; J64

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

†University College London. Email s.piyapromdee@ucl.ac.uk. I gratefully acknowledge the financial support from the Danish Council for Independent Research Social Sciences, grant DFF – 6109-00037

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

1 Introduction

What is the assignment of workers to firms, and how is it realized? 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. (2017) (BLM), our method identifies double sided unobserved heterogeneity through an application of the EM-algorithm. Using the estimated wages as the cardinal measure with which to rank types, the estimation allows a quantification of the contribution to sorting of each part of the mobility model. Furthermore, the estimate measures wage inequality and how sorting contributes to it over time.

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 worker. 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 as non-employment. 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.

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 is 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, 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

We start by developing a model for the data.

We use the matched employer-employee data from Denmark from 1985-2013. Wages are reported at annual frequency and mobility data of workers are reported at a weekly level. Our analysis panel starts in 1987 and we use information from 1985 and 1986 to distinguish between short and long tenure jobs in the stock of jobs in 1987. We restrict the sample to include only employment spells that start after individuals finished their highest levels of education. We remove any spells that start after the individuals turn 50 years old and treat any spells with less than an average of 25 work hours in a week as a non-employment spell.

Workers are indexed by $i \in \{1, ..., I\}$ and firms by $j \in \{0, 1, ..., J\}$, where $j = 0$ reflects non-employment. For each worker i, we observe a set of time-invariant characteristics z_i including gender and education. Firms also differ from each other by a set of observed, fixed characteristics ζ_j such as public/private status. Each worker i is either drawn from the working age population in 1987, or enters the panel in the first week of the first year following his or her last year of schooling. Individual trajectories $X_i = (w_{it}, j_{it}, x_{it})_{t=1}^{T}$ are recorded at weekly frequency, where $j_{it} \equiv j(i,t) \in \{0, 1, ..., J\}$ is the employer's ID in the t-th week of observation, x_{it} are observed worker controls including potential experience (age minus age upon leaving school), job tenure, and calendar time, and w_{it} is the worker's log-wage rate at occurrence t. Note that although the number of repeated observations T varies across individuals, we adopt the simplified notation of a balanced panel. Job mobility is measured at a weekly frequency but wages are measured at annual frequency. Thus, wage observations are missing except for the first week of any new match and the first week of the year.[1]

We assume that employers (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. Non-employment is observable and is denoted by $\ell = 0$.

We make two assumptions regarding worker and firm classifications. We omit the individual index i for simplicity and denote as $p(x)$ the probability mass of a random variable X, describing the distribution of some trait in the population of workers at a point x.

Assumption 1 (Initial condition)**.** *1) Initial x_1 does not predict worker type k conditional on z: $p(k|x_1, z) = p(k|z)$. 2) Initial employer type ℓ_1 is independent of z given k and x_1: $p(\ell_1|z, k, x_1) = p(\ell_1|k, x_1)$.*

These assumptions are natural. First, it is well known that the effects of time-invariant controls are not identified in fixed-effect models and are subsumed in the fixed effects. Second, by this way, we have a common scale of heterogeneity to compare workers of different gender or education and jobs in different industries. Third, there is *a priori* no loss of generality in proceeding this way. Suppose for example that wages are constant across all men and across all women, but differ across gender types. Then the best way of classifying wages given gender will be the align k on gender. However, in practice, K is likely to be lower that the number of different values of z. So there may be some loss of information.

Our next assumption is a conditional independence assumption giving a Markovian structure to the model.

Assumption 2 (Conditional independence)**.** *The next employer type and wage (w_t, ℓ_{t+1}) depend on the current information: z, $\boldsymbol{\ell}_t = (\ell_t, \ell_{t-1}, ...)$, $\mathbf{w}_t = (w_t, w_{t-1}, ...)$ and $\mathbf{x}_t = (x_t, x_{t-1}, ...)$ as follows*

$$p(w_t, \ell_{t+1}|z, k, \boldsymbol{\ell}_t, \mathbf{w}_t, \mathbf{x}_t) = p(w_t|k, \ell_t, x_t)\, p(\ell_{t+1}|k, \ell_t, x_t).$$

We implicitly assume that mobility to ℓ_{t+1} occurs at the end of period t, and hence is conditioned by x_t, like the wage w_t that is realized in period t.

The identification of finite mixture models is now beginning to be well understood (see e.g. Hall and Zhou, 2003; Hu, 2008; Hu and Schennach, 2008; Kasahara and Shimotsu, 2009; Allman et al., 2009; Hu and Shum, 2012; Henry et al., 2014; Bonhomme et al., 2016b,a;

[1] At the end of the year, employers declare to the tax administration services the cumulative salary paid to each of their employees during the elapsed year. Total salary divided by the total number of hours worked by the worker in that year is the wage rate that we assign to the first week of that year or of the next employment spell if the job started inside the year. Hours worked are inferred by Statistics Denmark from observed mandatory pension contributions that are conditional on weekly hours worked.

Gassiat et al., 2016). Recently, Bonhomme et al. (2017) have applied these identification tools to the a model of wages and mobility with two-sided heterogeneity, which can be understood as a non-parametric version of the model of Abowd et al. (1999) but with finite types. They show that two consecutive observations of wage and employer's type are generically sufficient to identify the model. In Appendix A we discuss two identification arguments that complements their analysis. First, we show how observed covariates can deliver identification with only one period. Second, we show identification with two periods if individuals can change employer within the same group of firms.

3 The estimation procedure

In this section we develop a Classification Expectation Maximization (CEM) algorithm for estimating the mixture model. We shall be treating the unobserved firm types $F = (\ell_1, ..., \ell_J)$ as fixed effects (i.e. a parameter to be estimated), and worker types $E = (k_1, ..., k_I)$ as random effects.

3.1 Likelihood

We state the likelihood for a given firm classification. Let $\ell_{it} = \ell_{j(i,t)}$ denote the type of the firm employing worker i in period t. Let also

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

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

For the given firm classification $F = (\ell_1, ..., \ell_J)$ let

$$q(\ell|\zeta, F) = \frac{\#\{j : \zeta_j = \zeta, \ell_j = \ell\}}{\#\{j : \zeta_i = \zeta\}} \quad \text{and} \quad q(\ell|F) = \frac{\#\{j : \ell_j = \ell\}}{J}$$

denote the share of type-ℓ firms given observed firm type ζ (e.g. industry and public/private status) and the unconditional share. We assume that conditional on drawing an employer of type ℓ, the probability of drawing this particular firm j with $\ell_j = \ell$ is inversely proportional to the frequency of this type, $q(\ell|F)$.

Let f denote a parametric version of the wage distribution $p(w|k, \ell, x)$; let M denote the transition probability $p(\ell'|k, \ell, x)$; let π denote the worker type probability $p(k|z)$; and let m denote the distribution of initial employer types $p(\ell_1|k, x_1)$. For a value $\beta = (f, M, \pi, m)$ of the parameters and a classification F of firms, the likelihood for one worker i – i.e. of

$X_i = (w_{it}, j_{it}, x_{it})_{t=1}^T$ conditional on (z_i, x_{i1}) – is

$$\sum_{k=1}^{K} L_i(k|\beta, F),$$

where the complete individual likelihood is

$$L_i(k|\beta, F) \equiv p(k, X_i|z_i, x_{i1}, \beta, F) = \frac{m(\ell_{i1}|k, x_{i1})\,\pi(k|z_i)}{q(\ell_{i1}|F)} \prod_{t=1}^{T} f(w_{it}|k, \ell_{it}, x_{it})$$
$$\times \prod_{t=1}^{T-1} M(\neg|k, \ell_{it}, x_{it})^{1-D_{it}} \left(\frac{M(\ell_{i,t+1}|k, \ell_{it}, x_{it})}{q(\ell_{i,t+1}|F)} \right)^{D_{it}}, \quad (1)$$

where $M(\neg|k, \ell, x) = 1 - \sum_{\ell'=0}^{L} M(\ell'|k, \ell, x)$ is the probability of staying with the same employer, and assuming that for the last observation period we do not know whether a mobility occurs or not by the end of it. By convention $f(w|k, \ell, x) = 1$ if the wage observation w is missing and $M(\ell'|k, \ell, x) = 0$ if $\ell = \ell' = 0$ (no transition from unemployment to unemployment). The term $m(\ell_{i1}|k, x_{i1})\,\pi(k|z_i)$ is the probability of the initial match type (k_i, ℓ_{i1}), for $k_i = k$, under Assumption 1. We assume that each firm within a group is equally likely to be selected. With that, the ratio $1/q(\ell_{it}|F)$ is proportional to the probability that *this particular* firm j_{it}, which is of type ℓ_{it}, be selected, either initially or upon job-to-job mobility. The rest of the likelihood factors in this way under Assumption 2.[2]

3.2 The EM algorithm for a given firm classification

The firm classification in the data is 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 co-dependency between a firm's workers resulting from their matches to a common firm type in a setup where workers move between firms. Consequently, we estimate the model for a given firm classification F. We shall explain in the next subsection how we set and update F.

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

$$p_i(k|\beta, F) \equiv \frac{L_i(k|\beta, F)}{\sum_{k=1}^{K} L_i(k|\beta, F)}. \quad (2)$$

[2] Note that the terms $1/q(\ell_{it}|F)$ do not show up in Bonhomme et al. (2017) because they do not imbed the estimation of the firm classification F in the same likelihood-maximization framework as the other parameters. Omitting them would make the Classification-step fail.

Note that the factors in one over $q(\ell|F)$ in the definition of $L_i(k|\beta, F)$ (equation (1)) appear in the numerator and the denominator of the definition of posterior probabilities in the same way and can be simplified out.

Then, define

$$Q_i(f|\beta^{(m)}, F) = \sum_{k=1}^{K} p_i(k|\beta^{(m)}, F) \left[\sum_{t=1}^{T} \ln f(w_{it}|k, \ell_{it}, x_{it}) \right]. \tag{3}$$

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

$$H_i(M|\beta^{(m)}, F) = \sum_{k=1}^{K} p_i(k|\beta^{(m)}, F) \left[\sum_{t=1}^{T-1} \left\{ (1-D_{it}) \ln M(\neg|k, \ell_{it}, x_{it}) + D_{it} \ln M(\ell_{i,t+1}|k, \ell_{it}, x_{it}) \right\} \right]. \tag{4}$$

It is the expected log-likelihood of worker i's employment history conditional on the first state ℓ_{i1}.

The EM algorithm iterates the following steps.

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

M-step Update $\beta^{(m)}$ by maximizing $\sum_i p_i(k|\beta^{(m)}, F) \ln L_i(k|\beta, F)$ subject to $\sum_k \pi(k|z) = 1$ for all z and $\sum_{\ell_1} m(\ell_1|k, x_1) = 1$ for all k, x_1, that is

$$f^{(m+1)} = \arg\max_f \sum_{i=1}^{I} Q_i(f|\beta^{(m)}, F), \tag{5}$$

$$M^{(m+1)} = \arg\max_M \sum_{i=1}^{I} H_i(M|\beta^{(m)}, F), \tag{6}$$

$$\pi^{(m+1)}(k|z) = \frac{\sum_{i=1}^{I} p_i(k|\beta^{(m)}, F)\, \mathbf{1}\{z_i = z\}}{\#\{i : z_i = z\}}, \tag{7}$$

$$m^{(m+1)}(\ell|k, x_1) = \frac{\sum_{i=1}^{I} p_i(k|\beta^{(m)}, F)\, \mathbf{1}\{x_{i1} = x_1, \ell_{i1} = \ell\}}{\sum_{i=1}^{I} p_i(k|\beta^{(m)}, F)\, \mathbf{1}\{x_{i1} = x_1\}}. \tag{8}$$

Note that $m(\ell_1|k, x_1)$ is identified for all levels of experience and tenure only in the first survey year. For all subsequent years, all workers entering the survey are also entering the labor market and have hence zero experience and tenure.

3.3 Firm re-classification

Given an initial value $\widehat{\beta}^{(s)}, F^{(s)}$, where $\widehat{\beta}^{(s)}$ can be obtained given $F^{(s)}$ using the previous EM algorithm, we update $F^{(s)}$ as

$$F^{(s+1)} = \arg\max_F \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k|\widehat{\beta}^{(s)}, F^{(s)}) \ln L_i(k; \widehat{\beta}^{(s)}, F).$$

In practice we only search for a firm reclassification increasing the likelihood. To do that we order firms by size from the largest to the smallest. We find $\ell_1^{(s+1)}$ keeping all other firm types equal to their values in $F^{(s)}$. Then we find $\ell_2^{(s+1)}$ given $\ell_1^{(s+1)}$ and $\ell_3^{(s)}, ..., \ell_J^{(s)}$. And so on until $\ell_J^{(s+1)}$. We then return to the EM iterations with the updated $F^{(s+1)}$. We call this algorithm a Classification EM algorithm as it resembles the eponym algorithm proposed by Celeux and Govaert (1992) as a variant of the EM algorithm of Dempster et al. (1977). See Appendix B for a detailed exposition of our CEM algorithm and why it is working.

This leaves the question of initialization of the firm classification, F^0. For this we opt for simplicity: We rank firms by average wage per worker in the firm, and cluster firms equally into J groups based on that sorting.

4 Empirical specification

In this section, we detail the chosen parametric specification of f and M.

4.1 Wage distribution

Wages are assumed lognormal given match type. Specifically,

$$f(w|k, \ell, x) = \frac{1}{\sigma_{k\ell}(x)} \varphi\left(\frac{w - \mu_{k\ell}(x)}{\sigma_{k\ell}(x)}\right), \tag{9}$$

with $\varphi(u) = (2\pi)^{-1/2} e^{-u^2/2}$. This specification of the log-wage mean allows for a match-specific mean $\mu_{k\ell}$ and variance $\sigma_{k\ell}^2$.[3]

The M-step update 5 takes the following form:

$$\begin{aligned}
\mu_{k\ell}^{(m+1)}(x) &= \frac{\sum_{i=1}^{I} p_i(k|\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} w_{it}}{\sum_{i=1}^{I} p_i(k|\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}} \\
[\sigma_{k\ell}^{(m+1)}(x)]^2 &= \frac{\sum_{i=1}^{I} p_i(k; |\beta) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} [w_{it} - \mu_{k\ell}^{(m+1)}(x)]^2}{\sum_{i=1}^{I} p_i(k|\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}}.
\end{aligned}$$

[3] The distribution f is the distribution of wages for matches with given characteristics. One may interpret these wages as productivity if there is no selection into employment conditional on types.

4.2 Transition probabilities

We omit conditioning on x_{it} to simplify the notations. The probability for a worker of type k of a transition from a firm of type $\ell = 1, ..., L$ to a firm of type $\ell' = 1, ..., L$ at time t is

$$M(\ell'|k, \ell, x) = \lambda_k(x)\nu_{\ell'}(x)P_{k\ell\ell'}(x). \tag{10}$$

Parameter $\lambda_k \in [0, 1]$ is the worker k conditional probability of a meeting with an outside employer. Parameter $\nu_{\ell'} \geq 0$, with $\sum_{\ell'=1}^{L} \nu_{\ell'} = 1$, is the probability that the outside draw is 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,

$$P_{k\ell\ell'} = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)}. \tag{11}$$

Parameter $\gamma_{k\ell}$, with $\sum_{\ell=1}^{L} \gamma_{k\ell} = 1$, measures the quality of the match (k, ℓ). If the worker draws a same-type job we assume that the worker moves with probability $1/2$.[4]

We also model unemployment-employment transitions in a completely unrestricted way:

$$M(\ell'|k, 0) = \psi_{k\ell'}, \quad M(0|k, \ell) = \delta_{k\ell}.$$

With this it follows that

$$M(\neg|k, 0) = 1 - \sum_{\ell'=1}^{L} M(\ell'|k, 0) = 1 - \sum_{\ell'=1}^{L} \psi_{k\ell'},$$

and for $\ell \geq 1$,

$$M(\neg|k, \ell) = 1 - \sum_{\ell'=0}^{L} M(\ell'|k, \ell) = 1 - \delta_{k\ell} - \lambda_{k\ell} \sum_{\ell'=1}^{L} \nu_{\ell'} P_{k\ell\ell'} = 1 - \delta_{k\ell} - \lambda_{k\ell} + \lambda_{k\ell} \sum_{\ell'=1}^{L} \nu_{\ell'}(1 - P_{k\ell\ell'}).$$

We prove in Appendix C that the more flexible specification $\lambda_{k\ell}, \nu_{k\ell'}, \gamma_{k\ell}$ is not identified from $M(\ell'|k, \ell) = \lambda_{k\ell}\nu_{k\ell'}P_{k\ell\ell'}$. In reality, for every choice of $M_{k\ell\ell'}$ for all k and $\ell, \ell' \geq 1$ there are two and only two solutions. Let $(\lambda_{k\ell}, \nu_{k\ell'}, \gamma_{k\ell})$ be such that

$$M(\ell'|k, \ell) = \lambda_{k\ell}\nu_{k\ell'} \frac{\gamma_{k\ell'}}{\gamma_{k\ell} + \gamma_{k\ell'}}.$$

[4] We experimented the specification $P_{k\ell\ell'} = \dfrac{\gamma_{k\ell'}}{\theta\gamma_{k\ell} + \gamma_{k\ell'}}$, 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.

Then we also have

$$M(\ell'|k,\ell) = \frac{\lambda_{k\ell}}{\gamma_{k\ell}}\,\nu_{k\ell'}\gamma_{k\ell'}\,\frac{1/\gamma_{k\ell'}}{1/\gamma_{k\ell}+1/\gamma_{k\ell'}}.$$

Hence $\left(\frac{\lambda_{k\ell}}{\gamma_{k\ell}}, \nu_{k\ell'}\gamma_{k\ell'}, \frac{1}{\gamma_{k\ell}}\right)$ is also a solution (with the appropriate normalizations). We thus need to restrict the parametrization further. We opt for making $\lambda_{k\ell}$ independent of ℓ and $\nu_{k\ell'}$ independent of k (but dependent on x). An employee will draw an alternative offer with probability λ_k, this offer is from a firm of type ℓ' with probability $\nu_{\ell'}$ and ℓ' beats ℓ with probability $P_{k\ell\ell'} = \frac{\gamma_{k\ell'}}{\gamma_{k\ell}+\gamma_{k\ell'}}$.

This parametric restriction on transition probabilities $M(\ell'|k,\ell,x)$ is a considerable reduction of dimensionality with respect to letting probabilities them unrestricted as in BLM. We believe that the loss of generality is amply compensated by the gains in efficiency and intelligibility. However, we also lose in simplicity. Transition probability estimates for the M-step of the EM algorithm are simple frequencies in the unrestricted case. Obtaining estimates in the the parametric restriction is another challenge. In Appendix D we develop an MM algorithm (Hunter, 2004; Hunter and Lange, 2004) that allows to maximize $H(M|\beta^{(m)})$ subject to the parametric restriction on M very rapidly.[5]

5 Estimation results

5.1 Data and estimation

The time-invariant worker characteristics z_i include gender and education. Education level is based on the normed number of years of education associated with the worker's highest completed degree. The low education group comprises all degrees normed to less than 12 years of education. The medium education group has a norm of exactly 12 years, and the high education group is any education level with a norm greater than 12 years. The time-invariant firm characteristics ζ_j include the public/private status. The worker's time variant characteristics x_i include the short/long tenure status, potential experience (time

[5]The MM algorithm works by finding a function that minorizes the objective function and that is more easily maximized. Let $f(\theta)$ be the objective concave function to be maximized. At the m step of the algorithm, the constructed function $g(\theta|\theta_m)$ will be called the minorized version of the objective function at θ_m if

$$g(\theta|\theta_m) \le f(\theta), \forall\theta, \text{ and } g(\theta_m|\theta_m) = f(\theta_m).$$

Then, maximize $g(\theta|\theta_m)$ instead of $f(\theta)$, and let $\theta_{m+1} = \arg\max_\theta g(\theta|\theta_m)$. The above iterative method guarantees that $f(\theta_m)$ converges to a local optimum or a saddle point as m goes to infinity because

$$f(\theta_{m+1}) \ge g(\theta_{m+1}|\theta_m) \ge g(\theta_m|\theta_m) = f(\theta_m).$$

since graduation), and calendar year.[6] Short tenure is defined as less than 100 weeks of employment. We divide experience into four groups: less than 5 years; 5-10 years; 11-15 years; and more than 15 years. Additionally, we allow the wage and mobility parameters to vary by 3 year time intervals. This leads to 9 different calendar time groups between 1987-2013.[7] Thus, x_{it} is one of 72 different groups.

Finally, after experimenting with different choices for K and L, we set the number of worker types to $K = 14$ and the number of firm types to $L = 24$. Worker and firm type assignments are assumed to be fixed over the duration of the panel. That is, tenure and experience impact a worker's mobility and wages conditional on a fixed type assignment. By implication, a type is characterized also by its dynamic wage and mobility paths. An often used alternative in the literature (see, for example Card et al. (2013)) is to stratify the data by calendar time intervals (a common length seems to be around 7 years) and estimate the model independently for each stratification.

The estimation groups workers with similar wage and mobility observations together. Similar for firms. We use wage observations to give a cardinal labeling to the groups. Specifically, we estimate $\mu_{k\ell}(x) = c(x) + a_k + b_\ell + \varepsilon_{k\ell}(x)$, then we relabel k and ℓ so that a_k and b_ℓ are now increasing in k, ℓ. Hence, by construction $\mu_{k\ell}$ will on average be increasing in the k and ℓ indices.

5.2 Distributions of worker and firm types

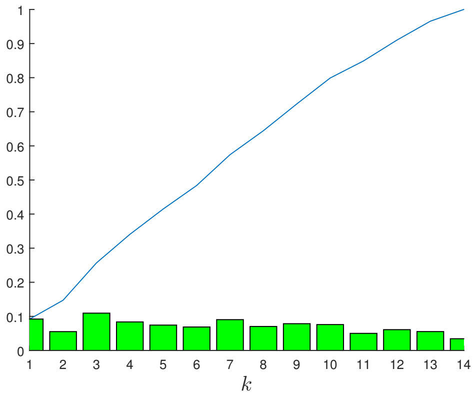
We calculate

$$p(k) = \frac{1}{N}\sum_i p_i(k), \quad p(\ell) = \frac{1}{NT}\sum_{i,t} \mathbf{1}\{\ell_{it} = \ell\},$$

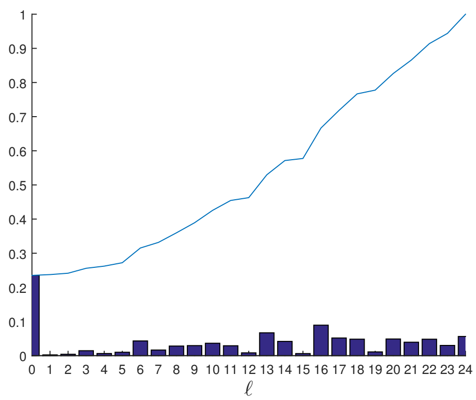
the marginal cross-sectional distributions of k and ℓ across workers. Figure 1 displays these distributions (plots (a) and (b)). The marginal CDFs of k and ℓ (we shall use $F(k), F(\ell)$ to denote them; the solid lines in Figure 1a,b) offer a different labeling for the plots that we are going to analyze. The advantage of using plotting for example $\mu_{k\ell}(x)$ with respect to $F(k), F(\ell)$ for a given x (year, tenure and experience) over using k and ℓ for axes is that we thus compress the axis-scale where there are few workers (as in the case of lower firm types, $\ell \leq 5$) and stretch it where there are many. Interestingly, some firm types are relatively rare in the population of firms (say $\ell = 11, 12, 13,$ 18-21$, 23, 24$) and frequent in the population of employed workers because these firms have larger sizes. The firm classification respects the assumption that firms within a group are equally likely to be sampled. Hence, to the extent

[6] We can calculate actual experience only for the workers entering the labor market after 1987. This is why we stick to potential experience, or age.

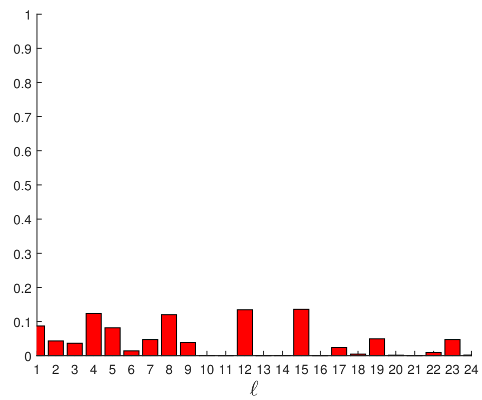
[7] We shall interpret calendar time as reflecting the macro environment. It could also refer to cohort effects. There is obviously no way to separate cohort, time and age.



(a) Marginal distribution of k, $p(k)$



(b) Distribution of ℓ across workers, $p(\ell)$



(c) Distribution of ℓ across firms, $q(\ell)$

Figure 1: Cross-sectional distributions of types

that firm size and vacancy posting are related (as one would expect), the firm classification will group by size, in addition to wage and mobility patterns.

Table 1 shows how observed characteristics z, gender and education, correlate with unobserved worker types (calculate mean posterior probabilities by observed type). There is no obvious correlation pattern. Some high types (large k) are mostly male, others mostly female, and again others half-half. The same conclusion applies to education. This is reassuring that observed characteristics do not predict well worker skill heterogeneity.

Table 2 repeats the operation for firms. Strikingly, most groups gather many firm that hire very few workers, and a few groups gather a small number of firms that hire many workers. For example, half of the 35 firms in group $\ell = 18$ and of the 45 firms of group $\ell = 11$ are public establishments and account for 8.8% and 3.1% of the 11.4 million spells in the panel. The 32 firms in the second largest group, $\ell = 13$, is mostly semi-public and account for 7.7% of all spells. Then the 74 firms in group $\ell = 21$, the 505 firms in group $\ell = 20$ and the 511 firms in group $\ell = 24$ are mostly private and account for 4.2%, 6.1% and

Table 1: Proportion of gender and education by worker type

Worker type	Gender		Education		
k	Male	Female	Low	Medium	High
1	0.15	0.85	0.46	0.47	0.07
2	0.47	0.53	0.48	0.37	0.15
3	0.59	0.41	0.62	0.33	0.05
4	0.33	0.67	0.27	0.56	0.17
5	0.30	0.70	0.24	0.53	0.23
6	0.38	0.62	0.15	0.43	0.41
7	0.76	0.24	0.22	0.66	0.11
8	0.59	0.41	0.17	0.59	0.25
9	0.61	0.39	0.12	0.52	0.36
10	0.74	0.26	0.20	0.55	0.24
11	0.53	0.47	0.05	0.26	0.69
12	0.71	0.29	0.06	0.35	0.59
13	0.78	0.22	0.04	0.25	0.71
14	0.83	0.17	0.04	0.20	0.76

6.5% of all spells. Although firm size and sector seem to be important factors to explain the estimated classification, it would be obviously difficult to explain why groups 11 and 18, and groups 20 and 24 are thus split.

Summing up, although we estimate a relatively small number of groups ($K = 14$ groups of workers and $L = 24$ groups of firms), this classification contains information that differs from the observed characteristics.

5.3 Wages

For conciseness, we summarize our results by presenting time-aggregated parameters since the qualitative features of these parameters are similar over time. In Figure 2, we show the time-aggregated $\mu_{k\ell}$ for one experience and tenure status (10-15 years of experience and more than 100 days of tenure). Wages are generally increasing in both tenure and experience, but apart from that the links between expected wages and match types by experience and tenure are similar and we do not show them (see the online appendix for exhaustive figures).

Wages vary significantly more with worker than firm types. It is furthermore notable that $\mu_{k\ell}$ is broadly speaking monotone in the ranked firm and worker types. Thus, wage ranks tend to be global. Worker types agree on the wage ranks of firms, and vice versa. Moreover the surface $(k, \ell) \mapsto \mu_{k\ell}$ is essentially flat except for the highest k and the lowest ℓ. These results are broadly consistent with the assumption of linear wage fixed effects in Abowd et al. (1999). We will later perform a variance decomposition exercise that will allow us to make this assertion more precise.

Table 2: Firm Characteristics

Firm type	Sector			Group size	Avg no. spells	%no. spells
ℓ	No cat.	Public	Private			not updated yet
1	0.14	0.10	0.76	31781	2	1.00%
2	0.07	0.05	0.88	15729	8	1.14%
3	0.09	0.04	0.87	13310	26	2.26%
4	0.16	0.05	0.79	45374	3	0.91%
5	0.07	0.06	0.87	29810	7	3.18%
6	0.15	0.11	0.74	5093	150	6.24%
7	0.03	0.07	0.91	17265	16	2.01%
8	0.36	0.05	0.58	43921	12	4.80%
9	0.08	0.13	0.79	14113	35	4.33%
10	0.05	0.76	0.18	251	1787	0.91%
11	0.18	0.55	0.27	49	7581	3.17%
12	0.04	0.06	0.90	49125	3	5.37%
13	0.67	0.14	0.19	36	23272	7.66%
14	0.39	0.15	0.46	216	3064	3.61%
15	0.20	0.04	0.76	49754	2	1.93%
16	0.17	0.54	0.29	35	30280	2.90%
17	0.06	0.15	0.79	8834	84	6.42%
18	0.06	0.10	0.83	1621	449	8.79%
19	0.03	0.04	0.93	18001	15	6.76%
20	0.07	0.15	0.78	492	1344	6.08%
21	0.11	0.08	0.80	61	7509	4.15%
22	0.10	0.10	0.80	3503	201	4.32%
23	0.16	0.06	0.78	17255	27	5.53%
24	0.07	0.11	0.82	589	1292	6.52%

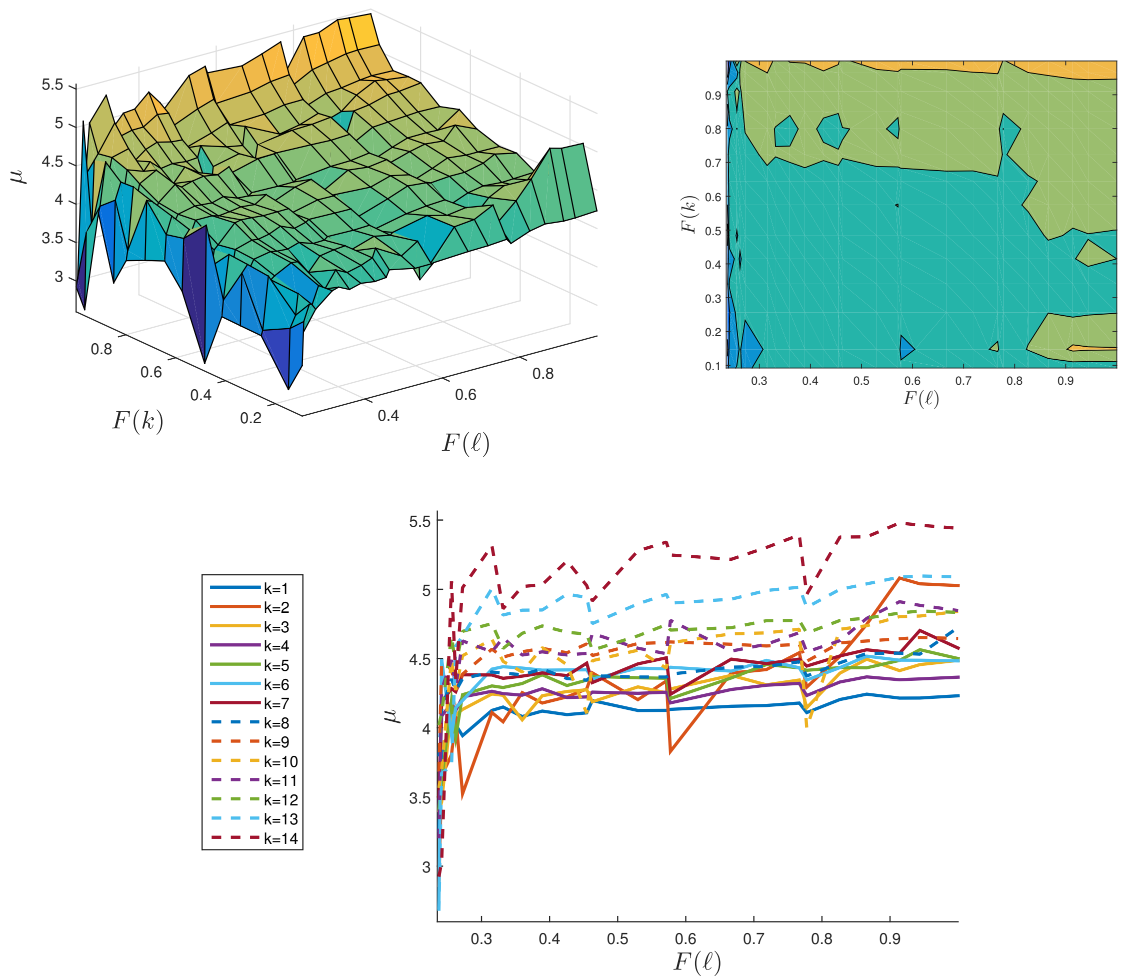


Figure 2: Mean wage $\mu_{k\ell}$ (10-15 years of experience, more than 100 days of tenure)

Table 3: Correlation with $\mu_{k\ell}$

	$\sigma_{k\ell}$	$\delta_{k\ell}$	$\psi_{k\ell}$	$\gamma_{k\ell}$	λ_k	$\nu_{\ell'}$
experience	tenure $<$ 100 weeks					
$<$5 years	-0.41	-0.81	0.50	0.32	0.42	0.16
5-10 years	-0.28	-0.73	0.54	0.28	0.50	0.18
10-15 years	-0.19	-0.64	0.56	0.25	0.47	0.22
$>$ 15 years	-0.16	-0.58	0.53	0.20	0.33	0.11
	tenure $>$ 100 weeks					
$<$5 years	0.10	-0.28	0.28	0.05	0.56	0.39
5-10 years	0.13	-0.29	0.19	-0.12	0.61	0.39
10-15 years	0.15	-0.26	0.16	-0.10	0.58	0.42
$>$ 15 years	0.14	-0.17	0.07	-0.17	0.35	0.41

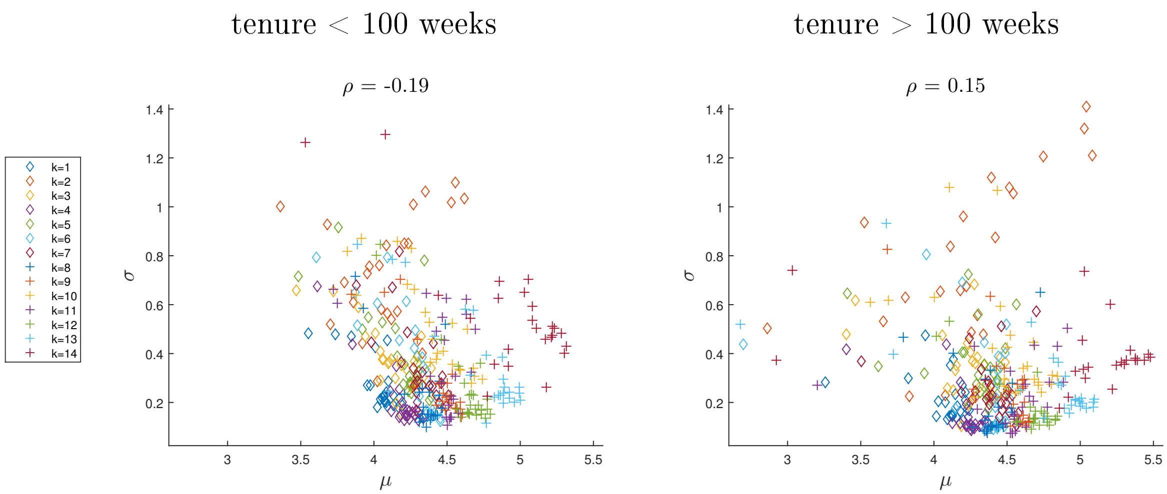
The correlation between idiosyncratic dispersion (as measured by standard deviation $\sigma_{k\ell}$) and mean wage $\mu_{k\ell}$ (with respect to the cross-sectional match distribution $p(k,\ell) = \frac{1}{NT}\sum_{i,t} p_i(k)\mathbf{1}\{\ell_{it} = \ell\}$) is generally negative and low. This negative link is stronger for less experienced and low tenure workers (see Table 3). However, idiosyncratic dispersion seems generally quite concentrated with a few outliers (mainly $k = 2$ and 3; see Figure 3).

5.4 Mobility

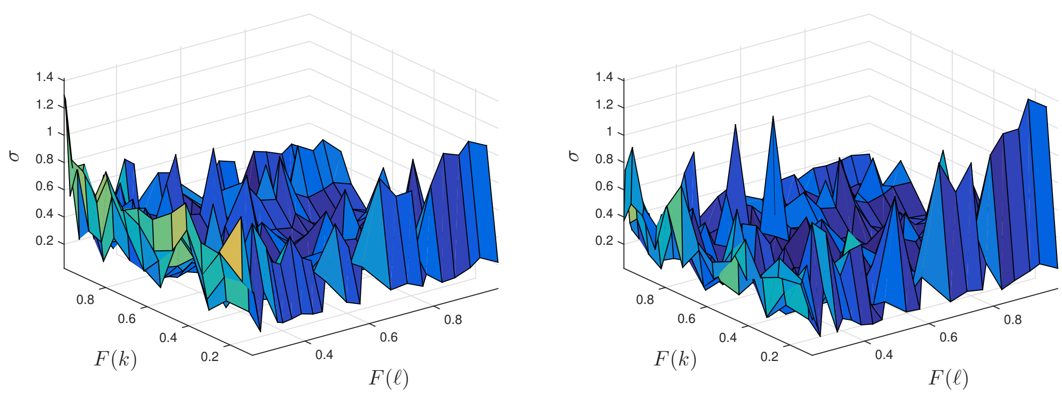
The top row of Figure 4 shows the search intensity parameter λ_k as a function the worker type-rank $F(k)$. Different lines correspond to different levels of experience. The left column plots search intensity for low tenure and the right column does it for high tenure. Overall, λ_k seems to be increasing in worker type k (with noise), and this relationship is much more pronounced for younger workers (the lines are ordered by decreasing experience and diverge). These plots also reveal that workers with low tenure search more intensely. It is worth noting that the core sorting mechanism in Bagger and Lentz (2013) implies that more skilled workers search more intensely, consistent with this finding.

The bottom row of Figure 4 shows the probability of drawing an offer from each firm type, $\nu_{\ell'}$. There is evidence that better firms are more likely to be sampled, and this relationship is stronger for more experienced and higher tenured workers (see also the correlations in Table 3).

Figures 5 and 6 depict the other mobility parameters $\delta_{k\ell}, \psi_{k\ell}, \gamma_{k\ell}$ by mean wage $\mu_{k\ell}$ and by worker and firm types. We only show the plots for workers with 10-15 years of experience as the different levels of experience do not change the general patterns. The strength of the link with worker and firm types varies though by experience. Table 3 displays the cross-sectional correlations with mean wage.

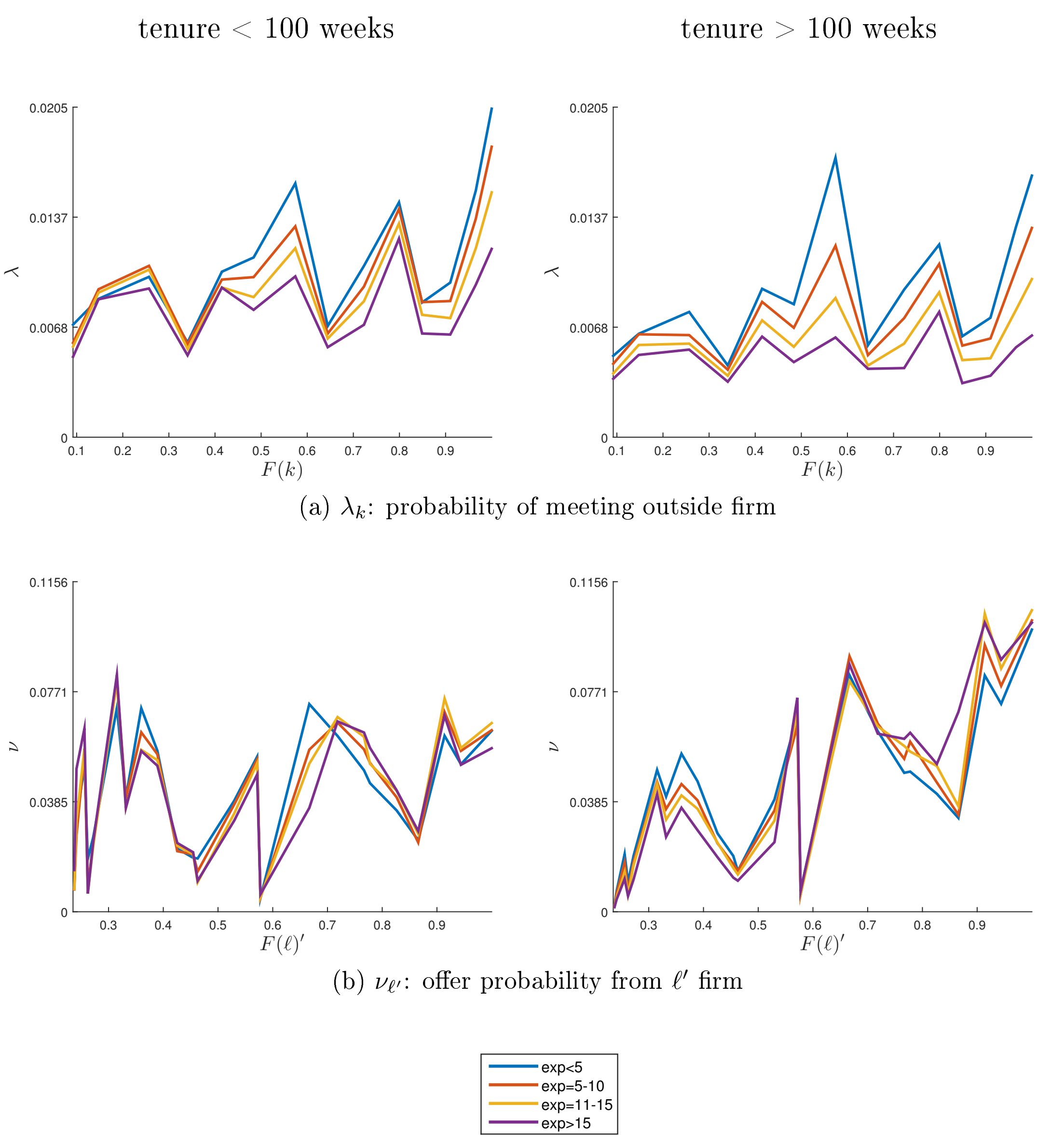


(a) By mean wage $\mu_{k\ell}$



(b) By worker and firm type-rank

Figure 3: Idiosyncratic dispersion $\sigma_{k\ell}$ (10-15 years of experience)



(a) λ_k: probability of meeting outside firm

(b) $\nu_{\ell'}$: offer probability from ℓ' firm

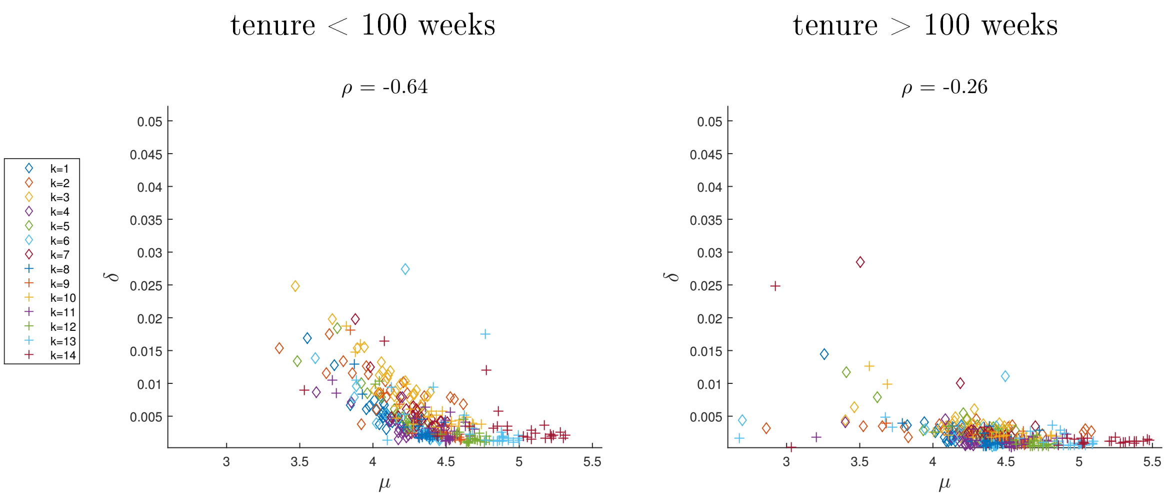
Figure 4: Search intensity λ_k and sampling probability $\nu_{\ell'}$

As can be seen, layoff rates $\delta_{k\ell}$ are strongly decreasing in mean wage, especially so for low tenure workers. The negative link between layoff rates and firm types is also emphasized in Bagger and Lentz (2013) and Jarosh (2015). Note, however, that for higher tenure workers at least the layoff rate is quite flat and homogeneous. For job finding rates $\psi_{k\ell}$ we find a positive link with mean wage, less strong than for layoff rates and again stronger for low tenure workers. Finally, the link between the parameter that measures the preference for the job (or match value), $\gamma_{k\ell}$, is nonexistent at longer tenure. It is weakly positive (correlation of the order of 0.3) at lower tenure. This may indicate that some workers, for example unemployed, may be forced to accept lower quality jobs that they will sooner than later quit when better opportunities show up. Now, the overall very low correlation between $\gamma_{k\ell}$ and $\mu_{k\ell}$ shows that the drivers of mobility across firm types are to a significant extent unrelated to wages. Many other factors such as job amenities, family shocks, and job specific idiosyncratic shocks may explain mobility in ways that is not necessarily optimal in terms of income.

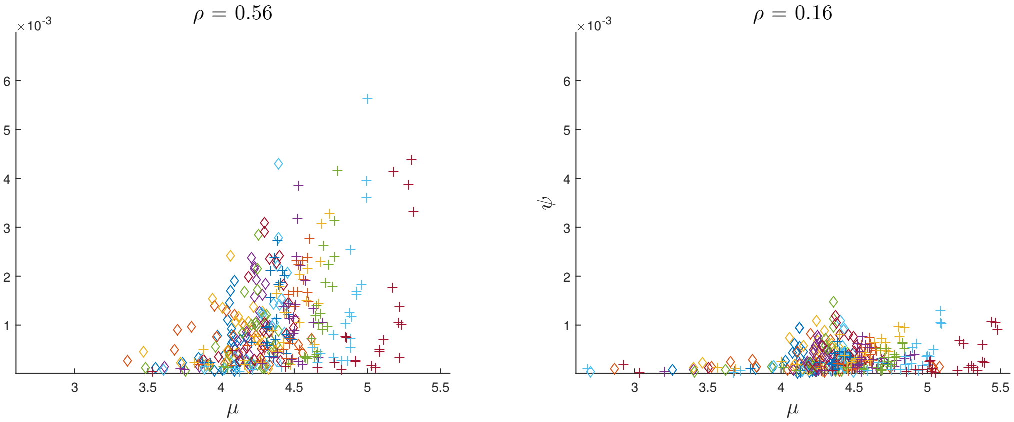
To explore the ladder structure further and the mobility implied by the λ, ν, γ estimates, Figure 7 presents the expected firm type destination conditional on the current firm type and a job-to-job move. For low skill workers (say $k \leq 7$; solid lines) the curves are essentially flat, except for experienced workers with long tenure. For workers of higher type ($k > 7$) then the expected destination firm rank is increasing in current firm rank, and more so for workers with short tenure. Now, in general, for all workers with tenure greater than 100 weeks or 2 years, the ladder effect is quite weak as these curves are quite flat. Most of the dispersion in expected destination firm rank is explained by worker unobserved heterogeneity. Therefore, as we concluded our analysis of $\gamma_{k\ell}$, we should not expect strong sorting resulting from better workers moving up the firm ladder faster.

6 Measuring sorting

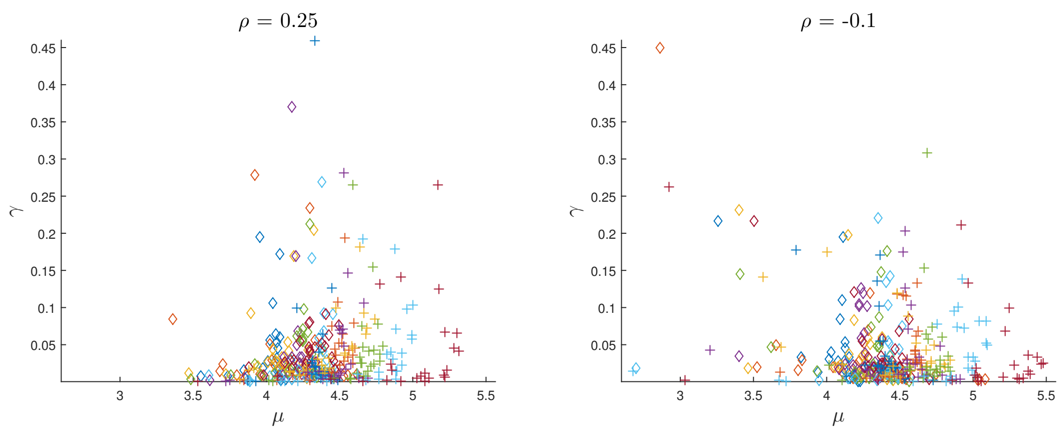
Both job-to-job moves as well as the mobility patterns in and out of unemployment point to positive sorting: Over time, job-to-job mobility implies that high type workers climb up the job ladder faster than low type workers leading to positive sorting, in a mechanism similar to that in Lise and Robin (2017); Bagger and Lentz (2013). More skilled workers are less likely to move into unemployment. When they do get unemployed, skilled workers move to higher ranked firms than less skilled workers do. However, this mechanism seems rather weak in reality. The aim of this section is first to provide a measurement of sorting, and second to understand how it is built.



(a) Layoff rate, $\delta_{k\ell}$

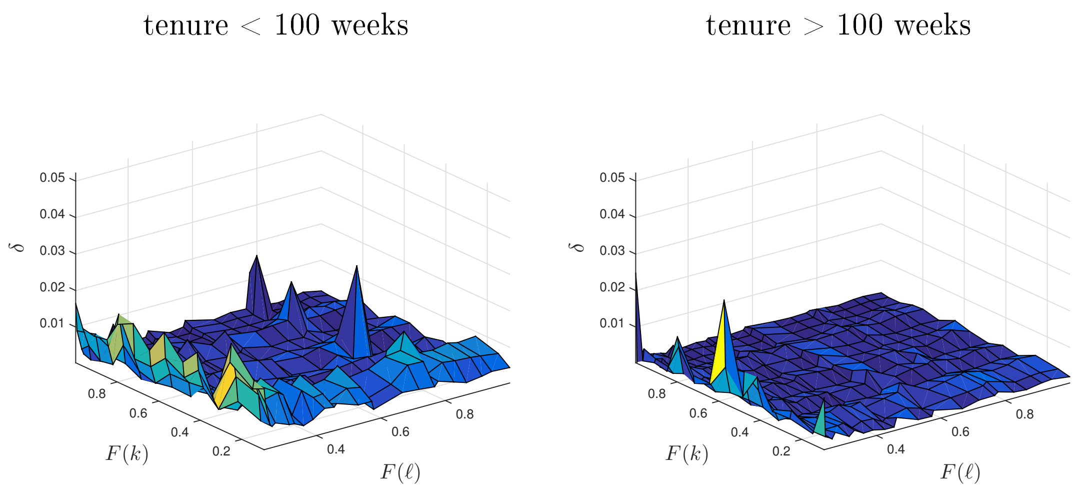


(b) Job finding rate for unemployed, $\psi_{k\ell}$

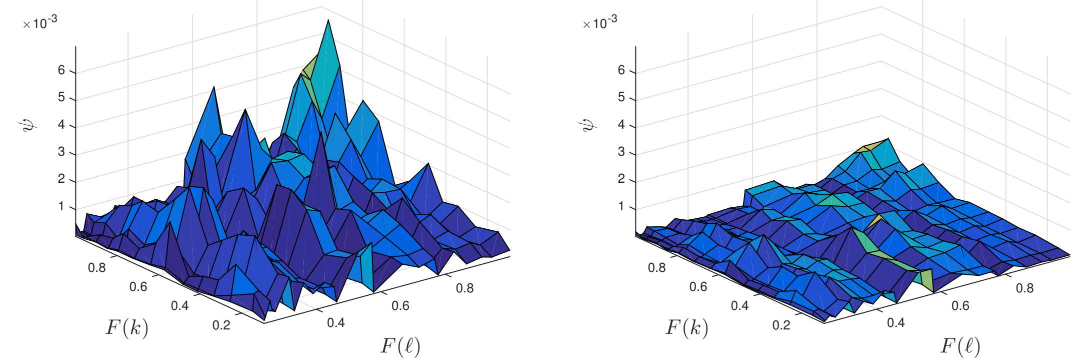


(c) Preference for the job, $\gamma_{k\ell}$

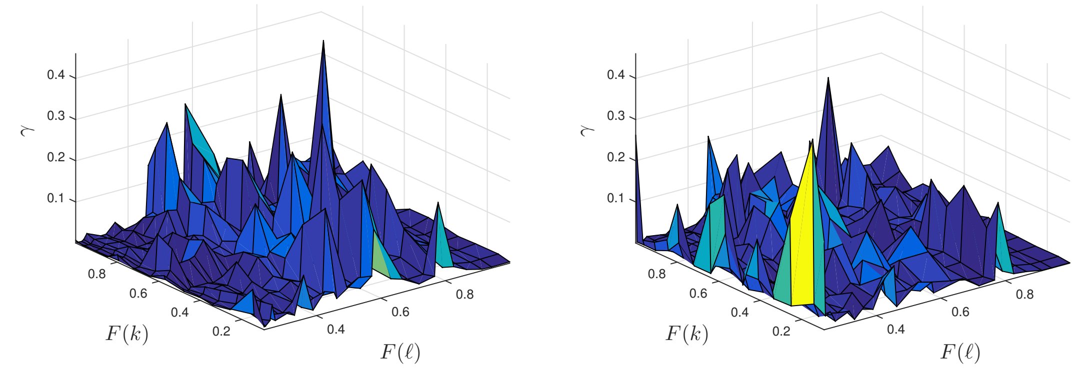
Figure 5: Mobility parameters by mean wage $\mu_{k\ell}$ (10-15 years of experience)



(a) Layoff rate, $\delta_{k\ell}$

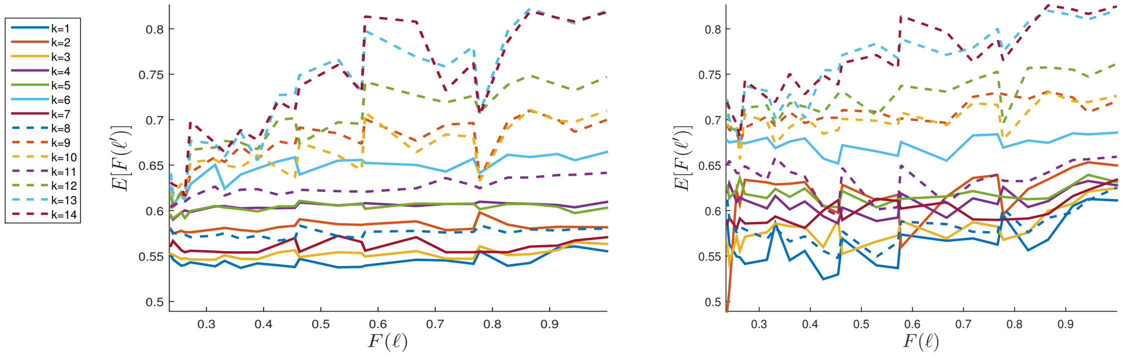


(b) Job finding rate for unemployed, $\psi_{k\ell}$

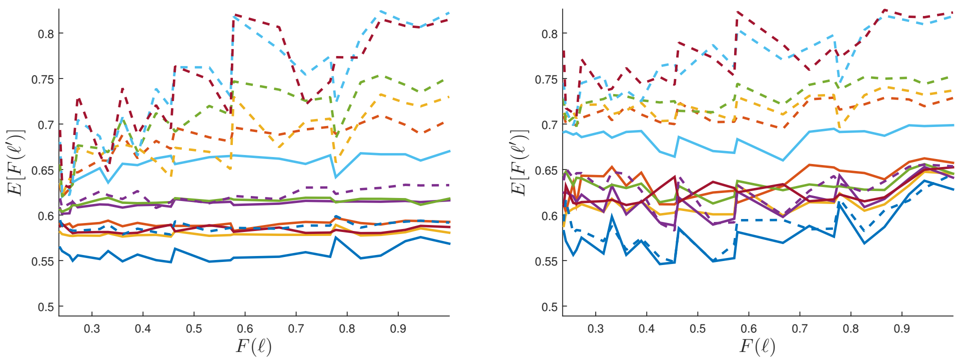


(c) Preference for the job, $\gamma_{k\ell}$

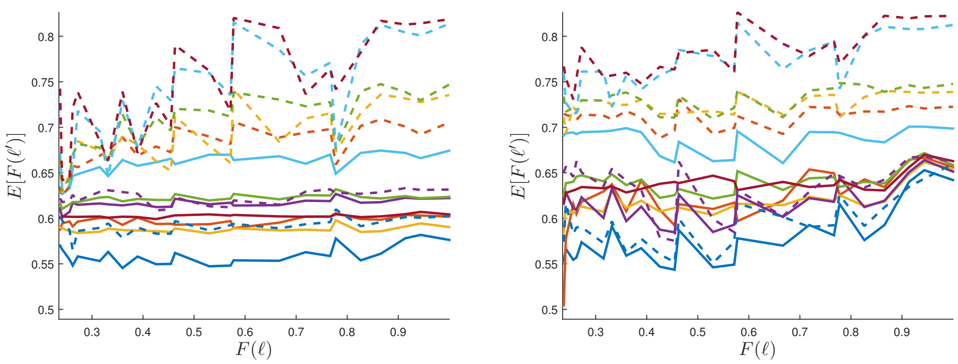
Figure 6: Mobility parameters by worker and firm type-rank (10-15 years of experience)



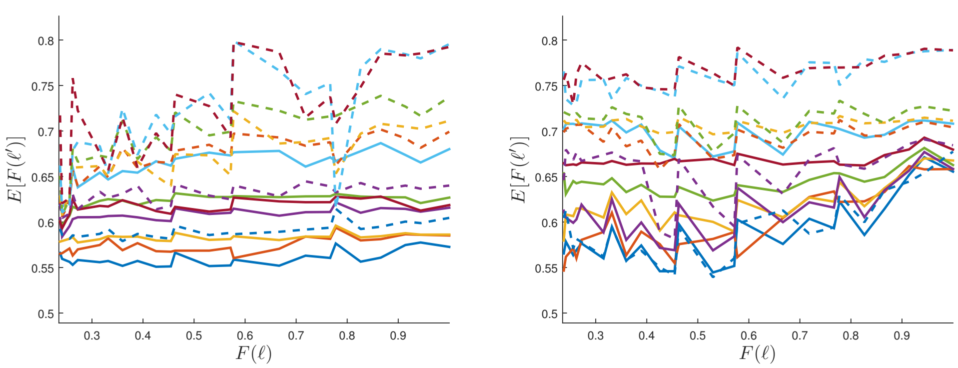
(a) Less than 5 years of experience



(b) 5-10 years of experience



(c) 10-15 years of experience



(d) at least 15 years of experience

Figure 7: Conditional mobility across firm types (short tenure: left, long tenure: right)

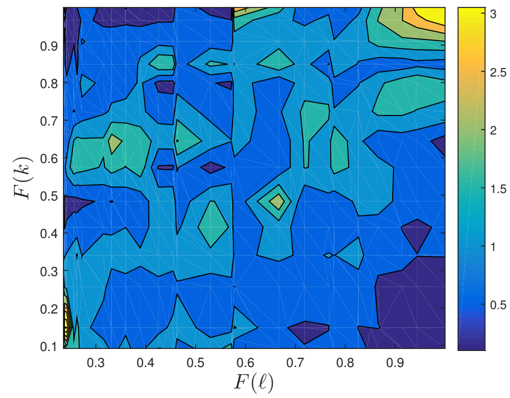


Figure 8: Copula $c[F(k), F(\ell)] \equiv \frac{p(k,\ell)}{p(k)p(\ell)}$

6.1 Distributions of matches

To illustrate the estimated sorting, Figure 8 shows the joint distribution of workers and firms relative to the matching probability under the assumption of independence, $c[F(k), F(\ell)] \equiv \frac{p(k,\ell)}{p(k)p(\ell)}$. This is a copula density. There is evidence of positive assortative matching. The entries in the first diagonal of the $c()$ mapping is significantly higher than off-diagonal entries. Low type workers tend to match with low type firms, high type workers with high type firms.

There is also significant noise in the relationship. To quantify the strength of the sorting, we show the correlation coefficient between the cardinal measures of worker and firm types we obtained in the wage projection. The worker type measure is a_k and the firm type measure is b_ℓ. We denote the correlation coefficient between a_k and b_ℓ in the estimated matching distribution by ρ_{ab}. Figure 9 shows the correlation coefficient between types conditional on time, experience, and tenure. The bottom panel shows the correlation coefficient conditional on time, only. As can be seen the correlation coefficients tend to take values between 0.25 and 0.4, confirming the positive sorting between firm and worker types. Furthermore, the positive sorting tends to get stronger from the beginning of the sample period until the early 2000's. The increasing pattern is particularly pronounced for long tenure matches. The findings based on AKM style wage fixed effect regressions in Bagger et al. (2013) and Card et al. (2013) suggest a stronger increasing sorting pattern over time than we find in our analysis. In particular, in the aggregate we find a fairly modest increase in the correlation coefficient over time from 0.25 to 0.3 (see Figure 10, line "Crossectional"). These estimates are considerably larger than those in Bagger et al. (2013) using the AKM model (from a low -.07 in 1981 to a high .14 in 2001). Overall, our sorting pattern is thus stronger and more stable.

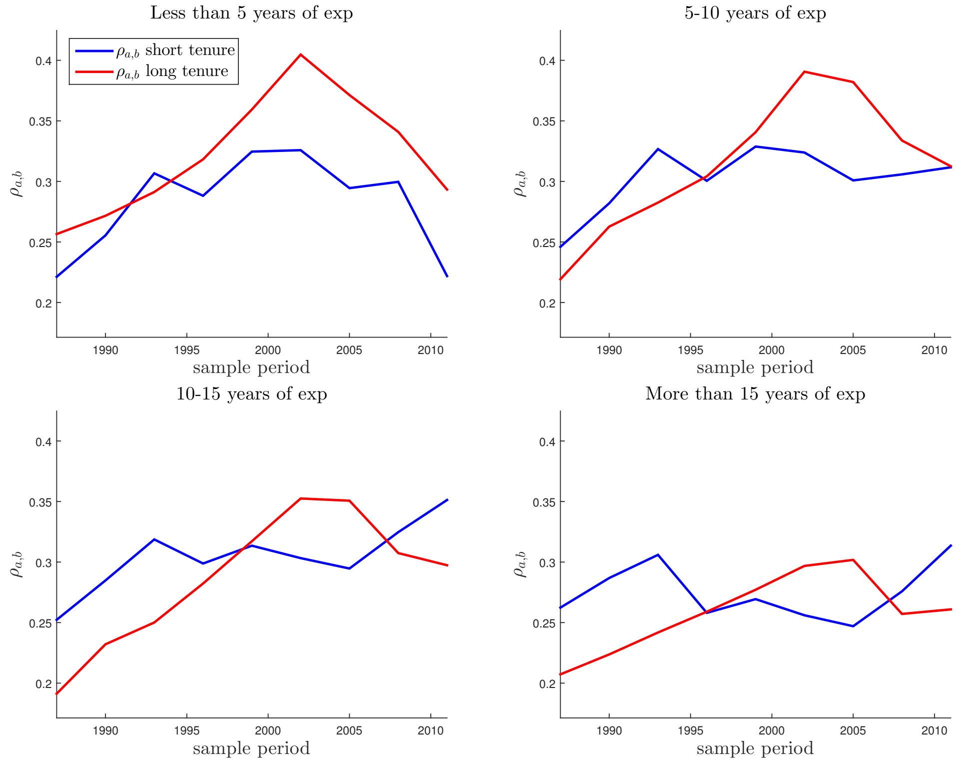


Figure 9: Sorting by tenure and experience

6.2 Mobility and sorting

Initial sorting and equilibrium sorting

Figure 10 compares three correlations between a_k and b_ℓ obtained from three different distributions: the observed cross-sectional match distribution $p(k,\ell)$ calculated for a given year; the steady-state (SS) distribution $p^\infty(k,\ell)$ that is obtained from the estimated transition probabilities $M_{k\ell\ell'}(x)$;[8] and lastly, the initial distribution of matches $p^0(k,\ell|x_1) = p(k)m(\ell|k,x_1)$, where $p(k) = \frac{1}{N}\sum_i p_i(k)$. We find that the $corr^0$ calculated with p^0 is above $corr$, calculated with p, which is above $corr^\infty$, calculated with p^∞. These correlations are close to each other and the differences may not be significant (although with so many observations, every point estimate tends to be very precisely estimated), but it is troubling to find that mobility is not a mechanism that effectively increases sorting. Mobility just helps maintaining sorting close to its initial level.

The anatomy of equilibrium sorting

To illustrate the relative strength of the different mobility channels that may impact sorting in the model, we perform a series of counterfactuals, summarized in Figures 11 and 12,

[8] From $M_{k\ell\ell'}(x)$ calculate $m^\infty(\ell|k,x)$ as the ergodic distribution associated with the Markov chain. Then, $p^\infty(k,\ell|x) = p(k)m^\infty(\ell|k,x)$, where $p(k) = \frac{1}{N}\sum_i p_i(k)$. Then average over x.

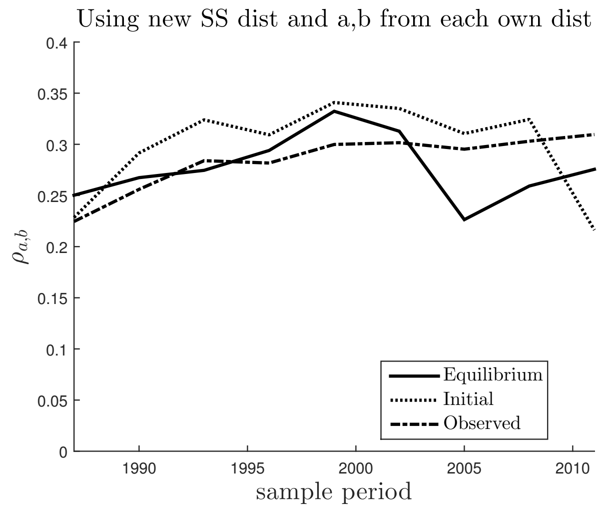
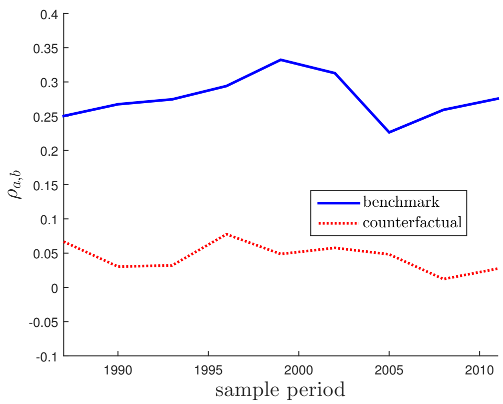


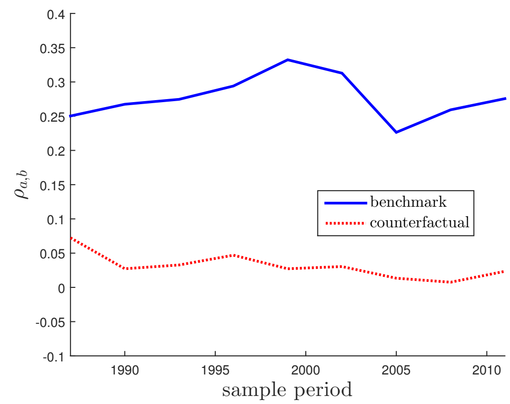
Figure 10: Observed, initial and equilibrium sorting

Table 4: Average counterfactual sorting over time

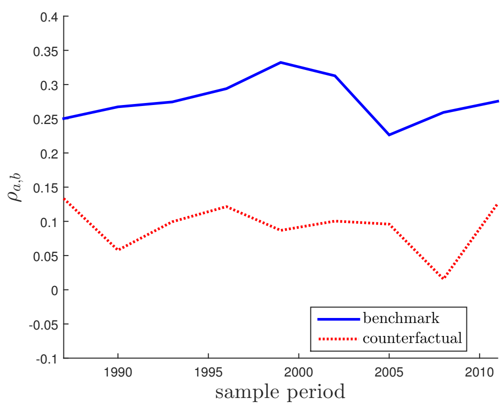
	Average $\rho_{a,b}$	% change
Benchmark	0.28	
Counterfactual		
No k variation in $\gamma_{k\ell}$	0.04	-83.92
No k,ℓ variation in $\gamma_{k\ell}$	0.03	-88.71
No k variation in $\psi_{k\ell}$	0.11	-59.67
No ℓ variation in $\psi_{k\ell}$	0.09	-66.42
No ℓ variation in ν_ℓ	0.06	-76.81
No k variation in λ_k	0.12	-56.76
No k variation in $\delta_{k\ell}$	0.16	-43.44
No ℓ variation in $\delta_{k\ell}$	0.14	-50.96
No k variations in $\delta_{k\ell}$ given no k variation in $\psi_{k\ell}$	0.11	-61.95
No tenure effect in $\gamma_{k\ell}$	0.15	-46.93
No tenure effect in λ_k	0.15	-46.50
No tenure effect in ν_ℓ	0.15	-47.12
No experience effect in $\gamma_{k\ell}$	0.16	-43.11
No experience effect in λ_k	0.15	-45.51
No experience effect in ν_ℓ	0.15	-45.94
No k variation in $\gamma_{k\ell}$ given no ℓ variation in ν_ℓ	0.00	-98.71



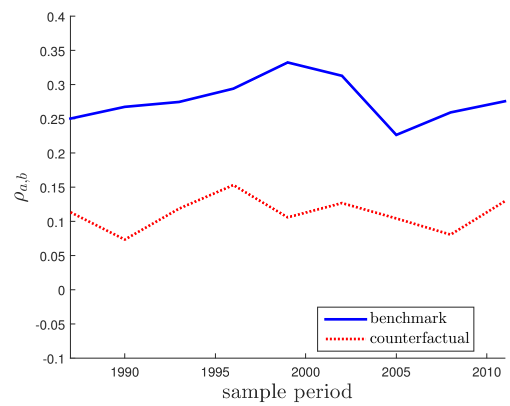
(a) $\gamma_{k\ell} = \overline{\gamma}_\ell$ (no k)



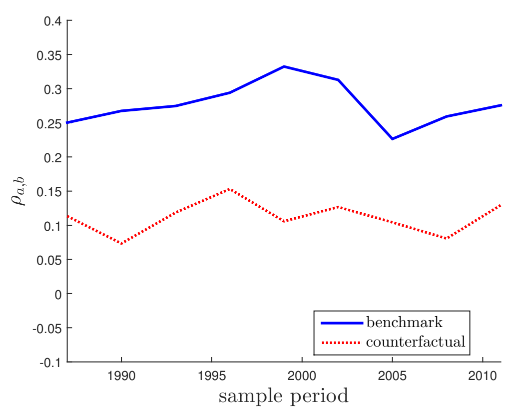
(b) $\gamma_{k\ell} = \overline{\gamma}$ (no k, no ℓ)



(c) $\psi_{k\ell} = \overline{\psi}_k$ (no ℓ)

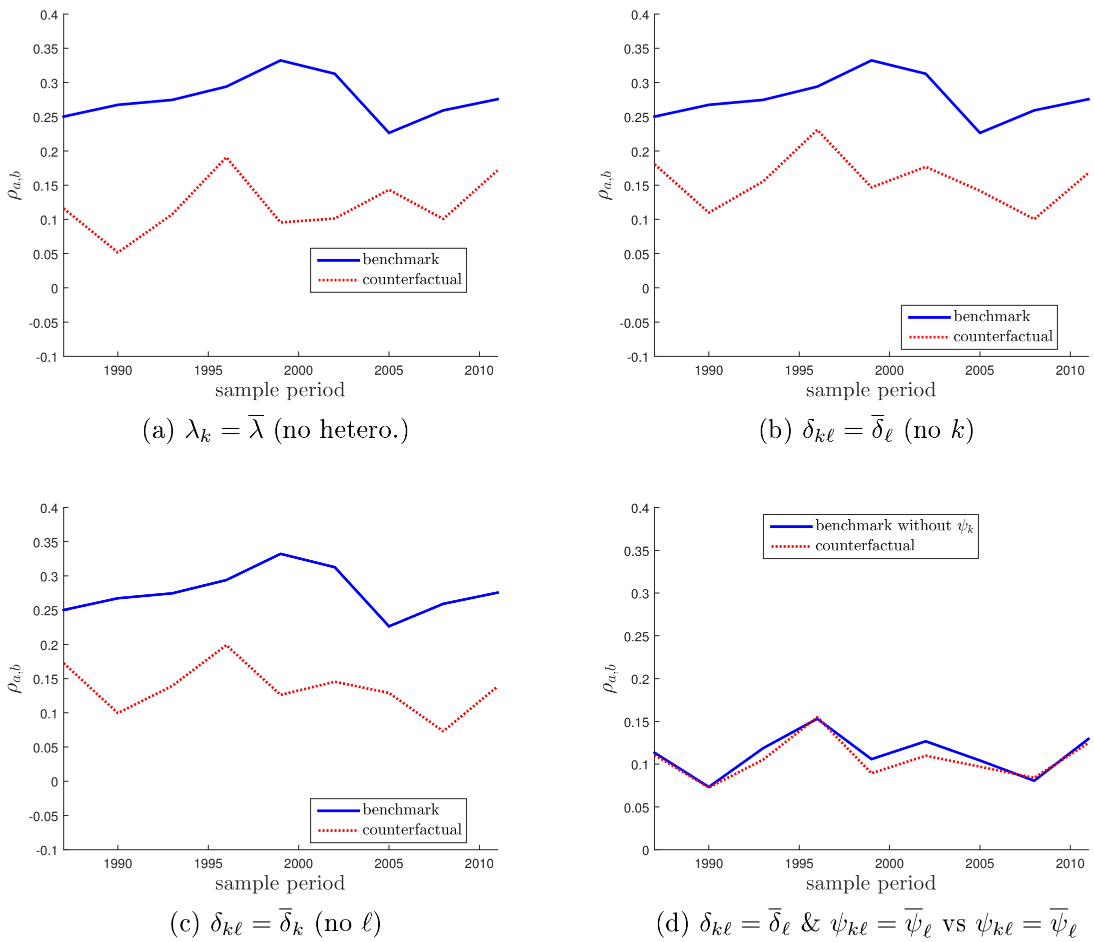


(d) $\psi_{k\ell} = \overline{\psi}_\ell$ (no k)



(e) $\nu_{\ell'} = \overline{\nu}$ (no hetero.)

Figure 11: Potent sources of sorting



(a) $\lambda_k = \overline{\lambda}$ (no hetero.)

(b) $\delta_{k\ell} = \overline{\delta}_\ell$ (no k)

(c) $\delta_{k\ell} = \overline{\delta}_k$ (no ℓ)

(d) $\delta_{k\ell} = \overline{\delta}_\ell$ & $\psi_{k\ell} = \overline{\psi}_\ell$ vs $\psi_{k\ell} = \overline{\psi}_\ell$

Figure 12: Ineffective sources of sorting

and in Table 4. Each counterfactual changes a particular mobility variable by removing the worker-type (or firm-type) variation and setting the variable to its worker-type (or firm-type) average in a given experience-tenure-time cell, while holding the others parameters constant. We then calculate the counterfactual steady state match probability $p^{\infty}_{alt}(k,\ell)$ and re-estimate the correlations between the cardinal measures of worker a_k and firm b_ℓ types in the wage projection as they have been estimated from the data.[9]

Figure 11a shows the counterfactual where $\gamma_{k\ell} = \overline{\gamma}_\ell$, i.e. there is no worker heterogeneity. If all workers agree on their preference for firms then the only sources of sorting are that better workers search more intensely and climb the common ladder faster, and that layoff rates may vary across worker types. This counterfactual removes all homophilic source of sorting. As can be seen, the impact on sorting is substantial. Figure 11b then considers removing firm heterogeneity from $\gamma_{k\ell}$. However, because of the normalizing constraint $\sum_\ell \gamma_{k\ell} = 1$, no firm heterogeneity means no heterogeneity at all, $\gamma_{k\ell} = \overline{\gamma}$. This explains why there is no substantial difference between the two counterfactuals.

Turning to the other significant source of sorting in the estimated model, Figure 11c shows the impact on sorting due to variation across worker types in the firm type destination distribution out of unemployment. Specifically, the counterfactual where $\psi_{k\ell} = \overline{\psi}_\ell$. We see this channel to be a significant source of sorting in the estimated model. It is a notable feature of the estimation since it is a channel that is typically absent in standard random search models. Figure 11d shows the counterfactual where we remove worker heterogeneity. The effect is the same, since the removal of either side of two-sided heterogeneity in a given source of sorting necessarily cancels out this channel as a potential source of sorting.

Figure 11e shows that the removal of heterogeneity in the sampling probability of firm types $\nu_{\ell'}$ has a significant effect of sorting, similar in magnitude to the effect obtained by removing heterogeneity in the job finding rates for unemployed. By contrast, Figure 12a demonstrates that offer arrival rate variation across worker types by itself has little impact on sorting. Hence an heterogeneous firm sampling probability does amplify sorting driven by heterogeneous preferences $\gamma_{k\ell}$, but not heterogeneous search intensity. This is likely due to the fact that search intensity does not vary enough across workers.

Figure 12b&c shows the impact on sorting from the layoff channel, $\delta_{k\ell} = \overline{\delta}_\ell$ and $\delta_{k\ell} = \overline{\delta}_k$. Even though low-quality matches are more likely to be sent into unemployment than high-quality matches, heterogeneous layoff risk does not significantly alter overall sorting. It is conceivable that the lack of impact on sorting could be a consequence of heterogeneous job finding rates for unemployed workers, $\psi_{k\ell}$. The estimated destination distribution out of unemployment goes a long way to directly place a worker type back to the position implied by the job-to-job mobility. Hence, layoffs could have minor sorting implications in our setting.

[9] We also tried with re-estimating $a_k$and b_ℓ with the counterfactual distribution. Results are similar.

Therefore, differential layoff differences across worker types would not affect sorting either. Figure 12d shows the impact of setting $\delta_{k\ell} = \overline{\delta}_\ell$ while also setting $\psi_{k\ell} = \overline{\psi}_\ell$. In this case, the comparison is with the counterfactual economy where $\psi_{k\ell} = \overline{\psi}_\ell$, only. As can be seen, also in this case, the cross worker type variation in layoff rates is not enough to significantly impact sorting.

7 Decomposing the variance of log-wages and wage growth

Finally, we measure the contribution of sorting to log-wage inequality and wage growth dispersion.

7.1 Log-wage levels

Variance decomposition formula

We are interested in decomposing the conditional variance of log wages w_{it} for a given calendar time period into person and firm effects. For each individual i present in the panel at any time, we can calculate $p_i(k)$ the posterior probability of being of type k using equation (2) at the estimated parameters (β and firm classification F), and for any given worker and employment-firm types k, ℓ and for any x indicating a particular tenure, experience and calendar time period, let $p(k, \ell, x) \propto \sum_{i,t} p_i(k) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}$ is the estimated proportion of observations with such k, ℓ, x values in the panel.

Next, for a given time period, let

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

be the linear projection of $\mu_{k\ell}(x)$ on all tenure and experience interactions and worker and firm indicators. The term $\widetilde{\mu}_{k\ell}(x)$ denotes the residual. That is, regress $\mu_{k\ell}(x)$ on tenure*experience dummies, worker dummies and firm dummies, weighing each (k, ℓ, x) by $p(k, \ell, x)$.

We can then decompose the log-wage variance as follows (omitting the conditioning on calendar time):

$$\begin{aligned}
\mathrm{V}\,(w_{it}) &= \mathrm{E}[V(w_{it}|k_i, \ell_{it}, x_{it})] + \mathrm{V}[\mathrm{E}(w_{it}|k_i, \ell_{it}, x_{it})] \\
&= \mathrm{E}[\sigma^2(k_i, \ell_{it}, x_{it})] + \mathrm{V}[\mu(k_i, \ell_{it}, x_{it})],
\end{aligned}$$

and

$$
\begin{aligned}
\mathrm{V}[\mu(k_i,\ell_{it},x_{it})] = \mathrm{V}[\overline{\mu}(x_{it})] + \mathrm{V}[a(k_i)] + \mathrm{V}[b(\ell_{it})] \\
+ \mathrm{V}[\widetilde{\mu}(k_i,\ell_{it},x_{it})] + 2\mathrm{Cov}[\overline{\mu}(x_{it}), a(k_i)] \\
+ 2\mathrm{Cov}[\overline{\mu}(x_{it}), b(\ell_{it})] + 2\mathrm{Cov}[a(k_i), b(\ell_{it})],
\end{aligned}
$$

where expectation operators (and variance and covariance) are with respect to distribution $p(k,\ell,x)$. For example,

$$
\mathrm{E}[\sigma^2(k_i,\ell_{it},x_{it})] = \frac{\sum_{k,\ell\neq 0,x} p(k,\ell,x)\,\sigma^2_{k\ell}(x)}{\sum_{k,\ell\neq 0,x} p(k,\ell,x)}.
$$

Results

Turning to Figure 13, we show the log-wage variance decomposition over time. In panel (a), it is seen that overall log wage variance, $\mathrm{V}(w)$, is increasing over time. Panel (b) shows the relative importance of within and between match wage variance, $\mathrm{EV}(w)$ and $\mathrm{VE}(w)$, respectively. As can be seen, the between match wage variance is increasing in importance over time so that by the end of the sample, in the early 2010s, the within match wage variance accounts for 60.5% of overall wage variance, from 43.8% at the end of the 1980s. The residual variance, the average of the idiosyncratic component $\sigma_{k\ell}(x)$, conversely decreased from around 56.2% to 39.5%. Note that we estimate considerably more idiosyncratic wage dispersion than is usually the case using the AKM model (see Table 5 for a comparison with Card et al., 2013; Song et al., 2015).[10]

Along with the increased importance of between match wage variance follows an increase in the variance across worker types $\mathrm{V}(a)$. Wage dispersion across worker types is the dominant source of dispersion across matches (around 26% between 1987 and 2007, and 35% after 2007), but its importance as a fraction of between match variation is decreasing over time (from 56.2% to 39.5%). Since we estimate so much more idiosyncratic variance, our estimates of person effect contribution is sizably smaller than in the AKM-based literature.

The effect of sorting $2\mathrm{Cov}(a,b)$ comes next, which explains about 7% of the overall log-wage variance. We confirm the increasing trend that has been observed elsewhere, although in a much more subdued way. The firm effect is comparable in size to the match-specific effect, around 5% (although recently they seem to diverge). This is evidence that worker and firm heterogeneity impact mean wages in a non additive way.

Summing up, we estimate much more residual variance than with the AKM model, and

[10] The mid-2000s seem to mark a structural change. This may be a data issue. CAN YOU CONFIRM THAT?

Table 5: Log-wage variance decomposition

	Card et al.		Song et al.		this paper				
	AKM		AKM		our estimator				AKM (no hetero.)
	85-91	02-09	80-86	07-13	87-89	02-04	11-13	avge	87-13
Person effect	61.3	51.2	47.5	52.8	28.7	24.8	36.0	29.1	44.7
Firm effect	18.5	21.2	16.0	11.9	4.9	6.1	3.8	4.8	17.1
Sorting	2.3	16.4	1.6	7.1	5.8	7.3	7.1	6.9	3.1
Observed hetero.	10.7	2.8	7.6	7.2	2.6	3.1	6.9	3.6	
Residual	8.2	5.9	20.3	14.9	52.1	53.3	38.5	50.1	35.1
Match effect					6.4	4.9	7.6	5.6	

Notes: Results for Song et al. (2015) are our own calculations from their Table 2. The worker effect is the sum of the "mean worker effect across firms" and the "difference of worker effect from mean worker effect across firms", and the effect of observed heterogeneity is the sum of the "mean xb across firms" and the "difference of xb from mean xb across firms". Results for Card et al. (2013) are extracted from their Table IV.

therefore a smaller contribution of worker effects. Sorting comes next and its contribution is increasing over the period. The firm effect and the match effect have similar, smaller contributions. The relative shares of worker, firm and sorting effects are consistent with the estimates in Bonhomme et al. (2017) using Swedish matched employer-employee data.[11] In order to both check the accuracy of our estimation procedure and evaluate the biases of AKM, we ran a Monte Carlo simulation as close as possible from the true data. The results can be found in Table 6. The simulations confirm a tendency of AKM to overestimate the contributions of worker and firm effects, and to underestimate the contribution of sorting as well as the residual variance.

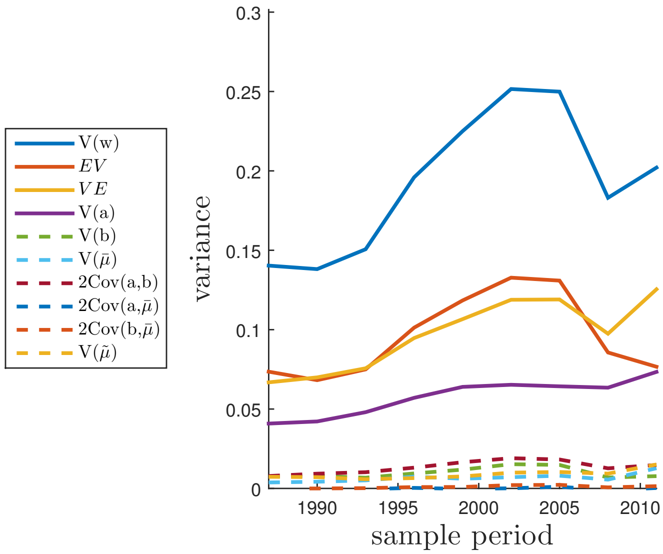
7.2 Wage growth upon job change

Variance decomposition formula

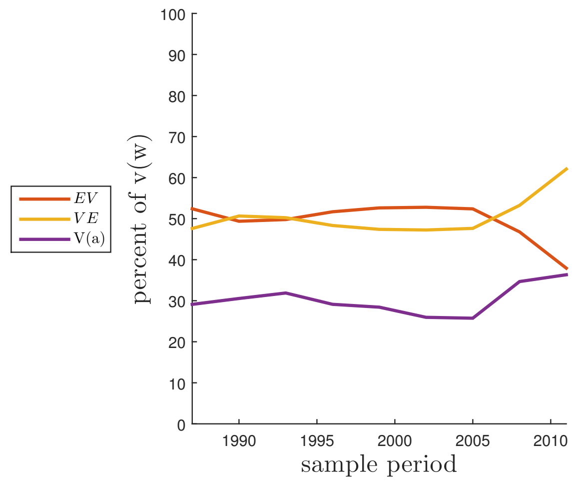
We can do a similar exercise for wage growth $\Delta \ln w_i = \ln w_{i2} - \ln w_{i1}$ conditional on being employed in two consecutive weeks 1 and 2 and changing employer. In period 1 the state would be some combination (k, ℓ, x). In period 2, she would have moved to a new firm of type $\ell' \in \{1, ..., L\}$ (maybe the same ℓ). Note that the new calendar time, tenure and experience triple x' is deterministic given ℓ, ℓ'.

The distribution of (k, ℓ, x) can still be described by $p(k, \ell, x)$. The probability of ℓ' given

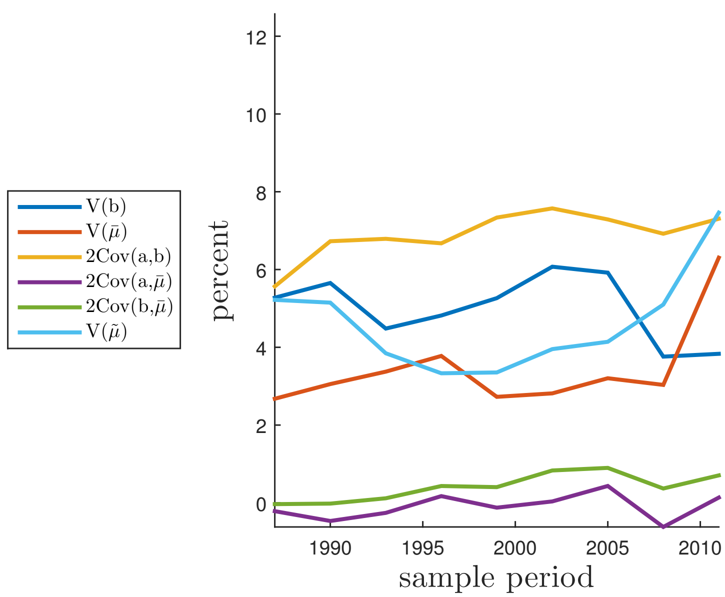
[11] We find relative shares of $\mathrm{V}(a)$, $\mathrm{V}(b)$ and $2\mathrm{Cov}(a, b)$ of about 68%, 13% and 18%. They find 80.3%, 3.4% and 16.3%. The contribution of firm effects is really small in their case.



(a) Levels



(b) Main components



(c) Firm effect and sorting

Figure 13: Variance decomposition

Table 6: Simulated log-wage variance decomposition

	Truth	AKM	LPR	Truth	AKM	LPR
	10-15 yrs exp, short tenure			10-15 yrs exp, long tenure		
Person effect	19.87	35.66	19.63	16.01	44.23	16.18
	(0.02)	(0.06)	(0.09)	(0.03)	(0.10)	(0.28)
Firm effect	6.04	10.29	6.11	25.19	22.97	25.65
	(0.01)	(0.04)	(0.04)	(0.08)	(0.12)	(0.39)
Sorting	7.17	4.19	7.07	6.25	-0.76	5.55
	(0.01)	(0.03)	(0.02)	(0.02)	(0.13)	(0.38)
Residual	63.62	49.87	64.00	45.24	33.56	45.78
	(0.04)	(0.05)	(0.05)	(0.05)	(0.07)	(0.12)
Match effect	3.27		3.19	7.31		6.83
	(0.00)		(0.04)	(0.02)		(0.12)

Notes: The simulations involves 1,000,000 workers and 100,000 firms where $K = 14, L = 24$. The length of the panel is 7 years. We run 100 simulations from the estimated model in steady state. The first column is the infeasible variance decomposition as if heterogeneity were observed. The second column refers to (therefore misspecified) AKM estimates. The third column refers to our model. For each variance component, the first row shows the mean value and the number in parenthesis the standard deviation across simulations.

(k, ℓ, x) and job change is $M(\ell'|k, \ell, x)$. Then

$$V(\Delta \ln w_i) = \mathrm{VE}(\Delta \ln w_i|k, \ell, x, \ell') + \mathrm{EV}(\Delta \ln w_i|k, \ell, x, \ell').$$

We have

$$\mathrm{EV}(\Delta \ln w_i|k, \ell, x, \ell') = \mathrm{E}\left[\sigma^2_{k\ell'}(x') + \sigma^2_{k\ell}(x)\right],$$

and

$$\begin{aligned}\mathrm{VE}(\Delta \ln w_i|k, \ell, x, \ell') &= \mathrm{V}\left[\mu_{k\ell'}(x') - \mu_{k\ell}(x)\right] \\ &= \mathrm{V}\left[\Delta\overline{\mu}(x) + \Delta a_k + \Delta b_{0\ell} + \Delta b_{1\ell\ell'} + \Delta\widetilde{\mu}_{k\ell\ell'}(x)\right],\end{aligned}$$

where the terms inside the last variance denote the linear projection of $\mu_{k\ell'}(x') - \mu_{k\ell}(x)$ on 1) tenure* experience dummies ($\Delta\overline{\mu}(x)$), 2) worker dummies (Δa_k), 3) employer dummies for movers ($\Delta b_{\ell\ell'}$), and 5) residuals ($\Delta\widetilde{\mu}_{k\ell\ell'}(x)$).

The following decomposition can finally be computed: 1) within: $\mathbb{E}\left(\sigma^2_{k\ell'}(x') + \sigma^2_{k\ell}(x)\right)$, 2) between x: $\mathrm{Var}\Delta\overline{\mu}(x)$, 3) between k: $\mathrm{Var}\Delta a_k$, 4) between ℓ, ℓ': $\mathrm{Var}\left(\Delta b_{\ell\ell'}\right)$, 5) match-specific effect: $\mathrm{Var}\Delta\widetilde{\mu}_{k\ell\ell'}(x)$, 6) sorting: $2\mathrm{Cov}\left(\Delta a_k, \Delta b_{\ell\ell'}\right)$, 7) all other covariances: $2\mathrm{Cov}\left(\Delta\overline{\mu}(x), \Delta a_k\right)$, $2\mathrm{Cov}\left(\Delta\overline{\mu}(x), \Delta b_{\ell\ell'}\right)$.

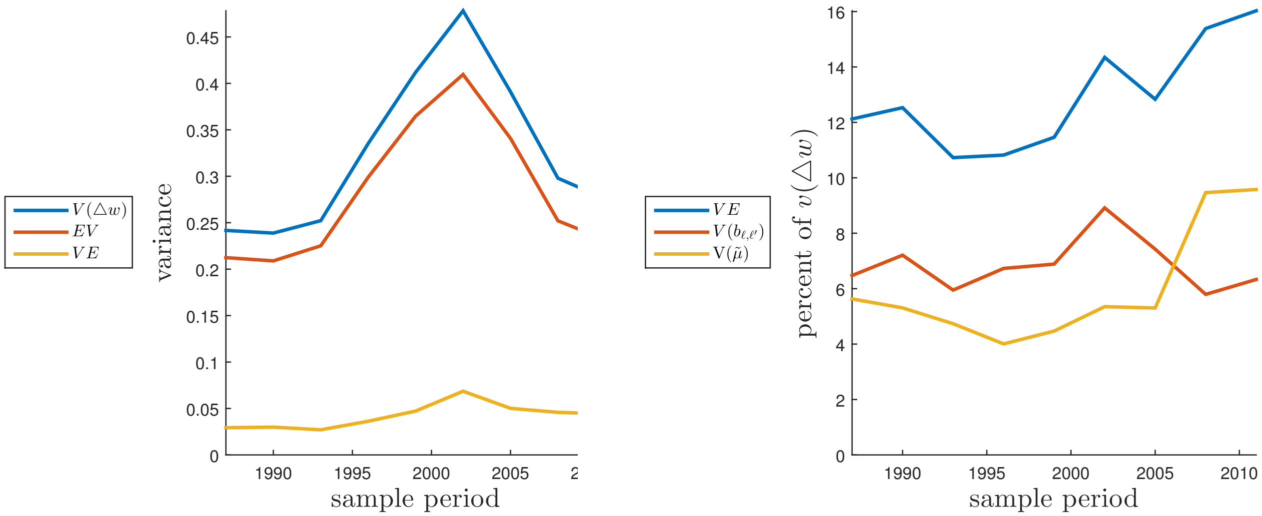


Figure 14: Variance decomposition of wage growth

Results

The results are displayed in Figure 14. Here most of the variance remains unexplained. The between variance is about 12% in 1987 and 16% in 2013. All components are negligible except for the match-specific effect and between-firm effects.

8 Conclusion

In this paper we use the finite mixture framework of Bonhomme et al. (2017) to estimate a model of wages and employment mobility on Danish panel of matched employer-employee data over the period 1987-2013. Our model allows for structural changes in the parameters and we propose a new parametrization for state transition probabilities. We develop a Classification Expectation Maximization algorithm allowing to estimate a random component model for $K = 14$ worker types and a classification of firms into $L = 24$ discrete groups. Our estimation algorithm works well and is fast, despite the nonlinear specification of state transition probabilities.

Our estimates of unobserved worker and firm heterogeneity and structural parameters shows an apparent disagreement in the way unobserved heterogeneity determines conditional mean wages, on one hand, and, on the other hand, the idiosyncratic wage variance and the mobility parameters. The strongest link is estimated for layoff rates and job finding probabilities for unemployed workers. The parameters governing the way a worker of a given type values job types display a much weaker link, and only at low tenure.

The joint distribution of match types shows evidence of moderate sorting. We measure sorting as the correlation of the worker and firm components obtained from a linear projection of conditional mean wages on worker and firm type dummies. Our estimates are around 25% and very stable over the all period.

We find that this correlation level is obtained at the first draw of a match in workers' careers. The moderate relationship between conditional mean wages and mobility parameters is only sufficient to maintain sorting at its initial level. There is no evidence of a strong ladder effect. The main parameters here are the preference for the job, the sampling distribution of firm types and the job finding rates of unemployed.

Finally, we estimate the log-wage decomposition. We estimate much more residual variance than with the AKM model, and therefore a smaller contribution of worker effects. Sorting comes next and its contribution is increasing over the period. The firm effect and the match effect have similar, smaller contributions. The relative shares of worker, firm and sorting effects are consistent with the estimates in Bonhomme et al. (2017) using Swedish matched employer-employee data.

APPENDIX

A Non-parametric identification

The basic linear algebraic tool for identification is the following lemma. Let there be two $m \times n$ matrices A_0 and A_1 such that

$$A_0 = G\Delta_0 F^\top, \quad A_1 = G\Delta_1 F^\top,$$

for two matrices G, F of dimensions $m \times k$ and $n \times k$, and two diagonal $k \times k$ matrices Δ_0, Δ_1. Suppose that $\text{rank}(A_0) = k \leq \min(m, n)$. This implies that G and F have linearly independent columns and that Δ_0 is invertible. Let the Singular Value Decomposition (SVD) of A_0 be $A_0 = U\Lambda V^\top$, for two orthogonal matrices U, V of dimensions $m \times k$ and $n \times k$, and an invertible $k \times k$ diagonal matrix Λ.[12]

Two interesting results follow. Firstly,

$$U^\top A_0 V\Lambda^{-1} = U^\top G\Delta_0 F^\top V\Lambda^{-1} = I_k \quad \text{(identity matrix)}.$$

This implies that if we let $W = U^\top G$, then $W^{-1} = \Delta_0 F^\top V\Lambda^{-1}$. Secondly,

$$U^\top A_1 V\Lambda^{-1} = U^\top G\Delta_1 F^\top V\Lambda^{-1} = W\Delta_1\Delta_0^{-1}W^{-1}.$$

This shows that $\Delta_1\Delta_0^{-1}$ is the diagonal matrix of the eigenvalues of $U^\top A_1 V\Lambda^{-1}$, and that W is one particular matrix of eigenvectors. If there are no multiple eigenvalues (all associated eigenspaces have dimension one), W is unique up to a multiplication of its columns by non zero numbers. If there are several such A_1 matrices, the second step's diagonalization becomes a simultaneous diagonalization problem.

A.1 Identification of remaining parameters with observed firm types

We first consider the case where the firm type is observed.

[12] The standard SVD has $A_0 = U\Lambda V^\top$ where U is $m \times m$ orthogonal, Λ is rectangular diagonal $m \times n$ and V is $n \times n$ orthogonal. If $\text{rank}(A_0) = k$, then we extract from U and V the first k columns, and from Λ the first k rows and columns (assuming this block contains the non zero singular values). Orthogonality of U and V means that $U^\top U = V^\top V = I_k$ (the identity matrix of size k).

A.1.1 One instrument, one wage, one job mobility

Assume that for each worker, we observe a set of characteristics z (gender, education, etc), the employer's type ℓ and the current wage w, as well as employment mobility at the end of the period (either no mobility, or mobility to a firm of type ℓ'). The worker type k is not observed. For simplicity, we neglect the control x and we assume that all variables are discrete.

Identifying matrices

We start by building identifying matrices using the following assumption specializing Assumptions 1 and 2 to the current case.

Assumption 3. *1)* $p(\ell|z,k) = p(\ell|k)$*; 2)* $p(w|z,k,\ell) = p(w|k,\ell)$*; 3)* $p(\ell'|z,k,\ell,w) = p(\ell'|k,\ell)$.

The first condition states that z helps measuring k but does not predict the employer's type ℓ given k. The second condition states the same for the wage w given the match type (k,ℓ) and z. The third condition says that job-to-job mobility only depends on the current match type (k,ℓ) and not on z or the current wage w.

Under Assumption 3,

$$p(k,z,\ell,w) = p(k)\,p(z|k)p(\ell|z,k)\,p(w|z,k,\ell) = p(k)\,p(z|k)\,p(\ell|k)\,p(w|k,\ell).$$

Then summing $p(k,z,\ell,w)$ over the unobserved worker type k,

$$p(z,\ell,w) = \sum_k p(z|k)\,p(k,\ell)\,p(w|k,\ell), \tag{12}$$

where $p(k,\ell) = p(k)\,p(\ell|k)$. Define the matrices

$$P_\ell = [p(z,\ell,w)]_{z\times w}, \quad G = [p(z|k)]_{z\times k}, \quad F_\ell = [p(w|k,\ell)]_{w\times k}, \quad \Delta_\ell = \mathrm{diag}[p(k,\ell)],$$

where the subscript indicates which variable indexes rows and which variable indexes columns. It follows from (12) that $P_\ell = G\Delta_\ell F_\ell^\top$.

Next, consider the probability of z, w, ℓ and of moving to a firm of any type ℓ' at the end of the period:

$$p(z,\ell,w,\ell') = \sum_k p(z|k)\,p(k,\ell)\,p(w|k,\ell)\,p(\ell'|k,\ell), \tag{13}$$

where the probability of moving to a new job of type ℓ' is $p(\ell'|z,k,\ell,w) = p(\ell'|k,\ell)$ by Assumption 3.

We can similarly define the matrices

$$P_{\ell\ell'} = [p(z, \ell, w, \ell')]_{z \times w}, \quad \Delta_{\ell\ell'} = \text{diag}[p(k, \ell, \ell')],$$

with $p(k, \ell, \ell') = p(k, \ell)\, p(\ell'|k, \ell)$. Equation (13) implies that $P_{\ell\ell'} = G\Delta_{\ell\ell'} F_{\ell'}^{\top}$.

Identification

We can then apply the lemma to our setup under the following assumption.

Assumption 4. *1) All matches form with positive probability: $p(k, \ell) \neq 0$ for all (k, ℓ). 2) $p(z|k)$ and $p(w|k, \ell)$ are linearly independent with respect to k given ℓ. 3) No two groups of workers k and k' have the same transition probabilities: $p(\ell'|k, \ell) \neq p(\ell'|k', \ell)$ for some (ℓ, ℓ'). 4) Workers of different type k have different distributions $p(z|k)$ of observed characteristics z.*

Condition 1 implies that matrix Δ_ℓ is non singular. It is a stronger assumption than necessary which is made to simplify the argument. We could allow for different sets of matching k's for each ℓ. Condition 2 guarantees that matrices G and F_ℓ are full-column rank. Both conditions together guaranty that the matrices P_ℓ are full column rank.

We can then apply the identification lemma with $A_0 = P_\ell$ and $A_1 = P_{\ell 1}, ..., P_{\ell L}$, separately for each initial employer type ℓ. Let $P_\ell = U_\ell \Lambda_\ell V_\ell^{\top}$ be the SVD of P_ℓ. The lemma shows that the matrices $U_\ell^{\top} P_{\ell\ell'} V_\ell \Lambda_\ell^{-1}, \ell' = 1, ..., L$, can be simultaneously diagonalized. The matrices of eigenvalues are $\Delta_{\ell\ell'} \Delta_\ell^{-1} = \text{diag}[p(\ell'|k, \ell)]$, which identifies transition probabilities. Under Condition 3 of Assumption 4 the entries of $\Delta_{\ell\ell'} \Delta_\ell^{-1}$ are all distinct. The common matrix of eigenvectors W_ℓ is such that

$$W_\ell = U_\ell^{\top} G D_\ell, \quad W_\ell^{-1} = D_\ell^{-1} \Delta_\ell F_\ell^{\top} V_\ell \Lambda_\ell^{-1},$$

for some non singular diagonal matrix D_ℓ. Hence, GD_ℓ is identified, and as the rows of G sum to one because $p(z|k)$ is a probability, it follows that the matrix D_ℓ is identified from the sum of the rows of $U_\ell W_\ell$. In the same way, $\Delta_\ell D_\ell^{-1}$ is also identified, hence Δ_ℓ, that is $p(k, \ell)$ for all k. The identification of G and F_ℓ immediately follows.

Finally, note that we have proceeded separately for workers employed in different sectors ℓ. This means that the classification of workers is contingent on the type of the first period's employer. One can identify worker groups across employer types ℓ if the distributions of observed types z vary across worker groups k (Assumption 4, Condition 4).

Wrapping up, if the distributions of observed worker characteristics and wages are sufficiently "rich" to differentiate worker types, and if the worker classification is minimal in the sense that no two groups are observationally identical and no group is degenerate (probability zero), then one wage observation and one job mobility are enough to identify the model when firm types are observed. If there are too few observed characteristics, say only gender, then the matrix G will not be full column rank for $K > 2$ and we shall need more information, such as two wage observations.

A.1.2 One job mobility, two wages

If there are no or insufficiently many observed worker characteristics, let us consider two consecutive periods of time, in which we observe the worker's wages and both employers' types. We now offer a simple proof of identification for this case that differs from BLM's in that we use both stayers and movers, and we consider the possibility of a mobility in the same job class (from ℓ to ℓ).[13]

Identifying matrices

We can proceed as in the previous subsection and build identifying matrices under a similar specialization of Assumptions 1 and 2.

Assumption 5. *1)* $p(\ell|z,k) = p(\ell|k)$*; 2)* $p(w|z,k,\ell) = p(w|k,\ell)$*; 3)* $p(\ell'|z,k,\ell,w) = p(\ell'|k,\ell)$*; 4)* $p(w'|z,k,\ell,w,\ell') = p(w'|k,\ell')$*.*

Assumption 5 adds to Assumption 3 the Condition 4, which is a conditional independence assumption stating that wages only depend on the current match type.

First, consider the probability of wages w and w' in periods 1 and 2 in the same firm of type ℓ (no job mobility is denoted as $\neg$):

$$p(\ell, w, \neg, w') = \sum_k p(k,\ell)\, p(w|k,\ell)\, p(\neg|k,\ell)\, p(w'|k,\ell),$$

where $p(\neg|k,\ell) = 1 - \sum_{\ell'} p(\ell'|k,\ell)$ is the probability of staying with the same employer given (k,ℓ). Let

$$F_\ell = [p(w|k,\ell)]_{w\times k}, \quad \overline{\Delta}_\ell = \mathrm{diag}[p(k,\ell,\neg)],$$

with $p(k,\ell,\neg) = p(k,\ell)\, p(\neg|k,\ell)$. We have $\overline{P}_\ell = [p(\ell,w,\neg,w')]_{w\times w'} = F_\ell \overline{\Delta}_\ell F_\ell^\top$.

Second, consider the probability of wages w and w' in periods 1 and 2 with a job mobility from ℓ to ℓ':

$$p(\ell, w, \ell', w') = \sum_k p(k,\ell)\, p(w|k,\ell)\, p(\ell'|k,\ell)\, p(w'|k,\ell').$$

[13] BLM's proof only uses job-to-job transitions and relies on the existence of some special, "alternating" job paths.

Denoting $\Delta_{\ell\ell'} = \text{diag}[p(k,\ell,\ell')]$, with $p(k,\ell,\ell') = p(\ell'|k,\ell)\, p(k,\ell)$, we have

$$P_{\ell\ell'} = [p(\ell,w,\ell',w')]_{w\times w'} = F_\ell \Delta_{\ell\ell'} F_{\ell'}^\top.$$

Identification

Now let us update Assumption 4 as follows.

Assumption 6. *1) $p(k,\ell)\neq 0$ for all (k,ℓ). 2) $p(w|k,\ell)$ is linearly independent with respect to k given ℓ. 3) For all (k,ℓ), no mobility is always possible: $p(\neg|k,\ell)\neq 0$. 4) For all (k,k') there exists ℓ such that $\frac{p(\ell|k,\ell)}{p(\neg|k,\ell)} \neq \frac{p(\ell|k',\ell)}{p(\neg|k',\ell)}$.*

Under Conditions 1 and 3, $\overline{\Delta}_\ell$ is non singular for all ℓ. Under Condition 2, F_ℓ is full-column rank. For the SVD $\overline{P}_\ell = U_\ell \Lambda_\ell U_\ell^\top$, let $W_\ell = \Lambda_\ell^{-1/2} U_\ell F_\ell \overline{\Delta}_\ell^{1/2}$. Then

$$W_\ell^{-1} = \overline{\Delta}_\ell^{1/2} F_\ell^\top U_\ell^\top \Lambda_\ell^{-1/2}.$$

It also holds that

$$\Lambda_\ell^{-1/2} U_\ell P_{\ell\ell} U_\ell^\top \Lambda_\ell^{-1/2} = W_\ell \overline{\Delta}_\ell^{-1} \Delta_{\ell\ell} W_\ell^{-1}.$$

Assuming that the entries of $\overline{\Delta}_\ell^{-1}\Delta_{\ell\ell} = \text{diag}[p(\ell|k,\ell)/p(\neg|k,\ell)]$ are all distinct (Condition 4), any matrix of eigenvectors is thus of the form

$$W_\ell = \Lambda_\ell^{-1/2} U_\ell F_\ell \overline{\Delta}_\ell^{1/2} D_\ell, \quad W_\ell^{-1} = D_\ell^{-1} \overline{\Delta}_\ell^{1/2} F_\ell^\top U_\ell^\top \Lambda_\ell^{-1/2},$$

for a non singular diagonal matrix D_ℓ. Since the rows of F_ℓ sum to one, then $\overline{\Delta}_\ell^{1/2} D_\ell$ and $\overline{\Delta}_\ell^{1/2} D_\ell^{-1}$ are both identified. Hence $\overline{\Delta}_\ell$ is identified, and so are D_ℓ and F_ℓ.

Now, having proceeded independently for each firm type ℓ, how do we know that the worker group k that we have thus labelled for firm type ℓ corresponds to the worker group k' that we have thus labelled for firm type ℓ'? Take matrix $P_{\ell\ell'}$ corresponding to a job mobility from ℓ to ℓ'. We know that $P_{\ell\ell'} = F_\ell \Delta_{\ell\ell'} F_{\ell'}^\top$ with $\Delta_{\ell\ell'}$ diagonal. So given arbitrarily chosen worker-group labels for ℓ and ℓ' one should relabel worker groups for ℓ' (say) by reordering the columns of $F_{\ell'}$ so that $F_\ell^{+} P_{\ell\ell'} F_{\ell'}^{+} = \Delta_{\ell\ell'}$ is a diagonal matrix (denoting $A^+ = (A^\top A)^{-1}A^\top$ for any full column rank matrix A).

Finally $\text{diag}[p(k,\ell)] = \overline{\Delta}_\ell + \sum_{\ell'} \Delta_{\ell\ell'}$ is also identified.

A.2 Identification of firm types from a cross-section of wages

Suppose that for each firm j we observe a set of characteristics ζ_j and two independent wages w_{j1}, w_{j2}. Moreover, suppose that for firms as for workers wages are independent of independent of observed types ζ_j given ℓ_j. Then, assuming independent wage draws given firm type, we can identify the proportion of each firm type, and the distributions of observed firm characteristics and wages given firm types.

More precisely let us make the following assumption. Let q denote probability distributions across firms.

Assumption 7. *1) Within a given firm of type (ℓ, ζ), the distribution of wages given (ℓ, ζ) is independent of ζ. 2) Wages are mutually independent given ℓ. 3) The distribution of ℓ is not degenerate ($q(\ell) \neq 0$ for all ℓ).*

Under this assumption, we can write the probability of (w_1, w_2) as

$$q(w_1, w_2) = \sum_{\ell} q(\ell)\, q(w_1|\ell)\, q(w_2|\ell)$$

and the probability of (w_1, w_2, ζ) as

$$q(w_1, w_2, \zeta) = \sum_{\ell} q(\ell, \zeta)\, q(w_1|\ell)\, q(w_2|\ell).$$

Per se ζ does not identify the firm classification. For example, suppose that we have manufacturing and services and public and private firms. We can classify firms into two groups by clustering them by their public/private status, or by industry. However, observable firm heterogeneity helps building matrices of moments with a common algebraic structure. Specifically, $P_\zeta = [q(w_1, w_2, \zeta)]$, w_1 in rows and w_2 in columns, can be factored as $P_\zeta = F \Delta_\zeta F^\top$, for $F = [q(w|\ell)]$ and $\Delta_\zeta = \text{diag}[q(\ell, \zeta)]$. Because $q(\ell) = \sum_\zeta q(\ell, \zeta) \neq 0$ by assumption, the identification lemma identifies $\Delta_\zeta \Delta_{\zeta_0}^{-1} = \text{diag}[q(\zeta|\ell)]$. Finally, using a similar argument as above, the matrix of eigenvectors, together with the fact that $q(w|\ell)$ is a probability and sums to one, identifies both conditional wage distributions F and type probabilities $q(\ell)$.

If we do not observe firm characteristics, a third wage observation is necessary to identify $q(\ell)$ and $q(w|\ell)$. And if we observe many wages per firm, then the posterior probability of firm type will precisely estimate the unobserved firm type. In this paper, we thus follow BLM's idea of classifying firms into distinct groups, using the employees' wages and other observed firm characteristics, while leaving worker types random. However, we shall iterate an algorithm that estimates the model first given a current firm classification, and then proceeds to an improved firm re-classification.

B The CEM principle

Given the statistical model which generates a set X of observed data, a set of two unobserved latent data or missing values E, F, and a vector of unknown parameters β, along with a likelihood function $p(X, E, F|\beta)$, the maximum likelihood estimate (MLE) of the unknown parameters is determined by the marginal likelihood of the observed data

$$p(X|\beta) = \sum_{E,F} p(X, E, F|\beta).$$

In our application E and F are two discrete variables: E is the set of worker types and F is the set of employer types (E for *employee* and F for *firm*). However, because different workers can be in the same firm at different periods of time, the summation over F can be extremely costly to compute and not even feasible. So instead we treat F as a parameter (a set of firm fixed effects) and we seek to maximize the conditional likelihood

$$p(X|F, \beta) = \sum_{E} p(X, E|F, \beta),$$

with respect to β and F. For this we use the following Classification Expectation Maximization (CEM) algorithm (not quite like the one in Celeux and Govaert, 1992, but related).

For a given value $F^{(s)}$, we run a standard EM algorithm treating $F^{(s)}$ as data. The EM algorithm consists of applying the following two steps iteratively:

Expectation step (E step): Calculate the expected value of the log likelihood function, with respect to the conditional distribution of E given $X, F^{(s)}$ under the current estimate of the parameters $\beta^{(m)}$:

$$Q(\beta|\beta^{(m)}, X, F^{(s)}) = \sum_{E} p(E|X, F^{(s)}, \beta^{(m)}) \ln p(X, E|\beta, F^{(s)}),$$

where

$$p(E|X, F, \beta) = \frac{p(X, E|F, \beta)}{\sum_E p(X, E|F, \beta)}.$$

Maximization (M step): Find

$$\beta^{(m+1)} = \arg\max_{\beta} Q(\beta|\beta^{(m)}, X, F^{(s)}).$$

The EM algorithm delivers a sequence of parameter values that increases the likelihood

$$p(X|F^{(s)}, \beta) = \sum_{E} p(X, E|F^{(s)}, \beta).$$

To see this, write the log likelihood of the data by,

$$\ln p(X|F,\beta) = \ln p(X,E|F,\beta) - \ln p(E|X,F,\beta).$$

Then multiplying both sides by $p(E|X,F^{(s)},\beta^{(m)})$ and summing over E,

$$\begin{aligned}\ln p(X|F^{(s)},\beta) = & \sum_E p(E|X,F^{(s)},\beta^{(m)}) \ln p(X,E|F^{(s)},\beta) \\ & - \sum_E p(E|X,F^{(s)},\beta^{(m)}) \ln p(E|X,F^{(s)},\beta) \\ = & \, Q(\beta|\beta^{(m)},X,F^{(s)}) + H(\beta|\beta^{(m)},X,F^{(s)}).\end{aligned}$$

By Gibbs' inequality $H(\beta|\beta^{(m)},X,F)] \geq H(\beta^{(m)}|\beta^{(m)},X,F)$. Hence, $Q(\beta|\beta^{(m)},X,F^{(s)}) + H(\beta^{(m)}|\beta^{(m)},X,F^{(s)})$ is a minorization of $\ln p(X|F^{(s)},\beta)$ in the point $\beta^{(m)}$, and improving $Q(\beta|\beta^{(m)},X,F^{(s)})$ over $Q(\beta^{(m)}|\beta^{(m)},X,F^{(s)})$ improves $p(X,F^{(s)}|\beta)$ relative to $p(X,F^{(s)}|\beta^{(m)})$.

This is well known. Let $\widehat{\beta}^{(s)}$ denote the estimate obtained from $F^{(s)}$. One can let the EM algorithm increase the likelihood $L(\beta,F^{(s)}|X)$ until convergence or stop after any number of iterations. We then update $F^{(s)}$ as follows.

Classification step (C step): Find

$$F^{(s+1)} = \arg\max \sum_E p(E|X,F^{(s)},\widehat{\beta}^{(s)}) \ln p(X,E|F,\widehat{\beta}^{(s)}).$$

By the same argument as before

$$\ln p(X|F,\beta) = \sum_E p(E|X,F^{(s)},\beta) \ln p(X,E|F,\beta) - \sum_E p(E|X,F^{(s)},\beta) \ln p(E|X,F,\beta).$$

Therefore, improving $\sum_E p(E|X,F^{(s)},\widehat{\beta}^{(s)}) \ln p(X,E|F,\widehat{\beta}^{(s)})$ delivers $F^{(s+1)}$ such that

$$p(X|F^{(s+1)},\widehat{\beta}^{(s)}) > p(X|F^{(s)},\widehat{\beta}^{(s)}) > p(X|F^{(s)},\widehat{\beta}^{(s-1)}).$$

The sequence $(\beta^{(s)},F^{(s)})$ delivers an increasing sequence of $(p(X|F^{(s)},\beta^{(s)})$ that converges to a local maximum of $p(X|F,\beta)$. This CEM algorithm is essentially like a sequential EM algorithm where β is updated given F and then F given β.

C Parametric identification of transition probabilities

We show that the parameters $\lambda_{k\ell}, \nu_{k\ell'}, \gamma_{k\ell}$ are identified given $M_{k\ell\ell'} = \lambda_{k\ell}\nu_{k\ell'}\frac{\gamma_{k\ell'}}{\gamma_{k\ell}+\gamma_{k\ell'}}$ for $\ell, \ell' = 1, ..., K$. Let us omit the index k to simplify the notation.

We have $M_{\ell\ell} = \frac{1}{2}\lambda_\ell \nu_\ell$, which identifies λ_ℓ given ν_ℓ. Moreover,

$$\frac{M_{\ell\ell'}}{2M_{\ell\ell}} = \frac{\nu_{\ell'}}{\nu_\ell}\frac{\gamma_{\ell'}}{\gamma_\ell + \gamma_{\ell'}}, \quad \frac{M_{\ell'\ell}}{2M_{\ell'\ell'}} = \frac{\nu_\ell}{\nu_{\ell'}}\frac{\gamma_\ell}{\gamma_\ell + \gamma_{\ell'}},$$

which identifies the parameter ratios $\frac{\gamma_\ell}{\gamma_{\ell'}}$ and $\frac{\nu_\ell}{\nu_{\ell'}}$ from the two ratios of transition probabilities $\frac{M_{\ell\ell'}}{2M_{\ell\ell}}$ and $\frac{M_{\ell'\ell}}{2M_{\ell'\ell'}}$. Note that

$$r_{\ell\ell'} \equiv \frac{1 - \sqrt{1 - \frac{M_{\ell\ell'}}{M_{\ell\ell}}\frac{M_{\ell'\ell}}{M_{\ell'\ell'}}}}{1 + \sqrt{1 - \frac{M_{\ell\ell'}}{M_{\ell\ell}}\frac{M_{\ell'\ell}}{M_{\ell'\ell'}}}} = \frac{\gamma_\ell + \gamma_{\ell'} - |\gamma_\ell - \gamma_{\ell'}|}{\gamma_\ell + \gamma_{\ell'} + |\gamma_\ell - \gamma_{\ell'}|} = \frac{\min\{\gamma_\ell, \gamma_{\ell'}\}}{\max\{\gamma_\ell, \gamma_{\ell'}\}}.$$

We would like to know which of $\gamma_\ell, \gamma_{\ell'}$ is the min and which is the max. If we knew the ordering of (γ_ℓ) then we could transform $r_{\ell\ell'}$ so as $r_{\ell\ell'} = \frac{\gamma_\ell}{\gamma_{\ell'}}$, and the normalization $\sum_\ell \gamma_\ell = 1$ would identify each γ_ℓ precisely. Unfortunately, we do not know a priori the ordering of (γ_ℓ).

Note first that $r_{\ell\ell'} = \frac{\gamma_\ell}{\gamma_{\ell'}}$ for all ℓ, ℓ' implies that $r_{\ell'\ell''} = r_{\ell\ell''}/r_{\ell\ell'}$ for all ℓ, ℓ', ℓ''. However, suppose that $\gamma_2 < \gamma_1 < \gamma_3$. Then $r_{12} = \frac{\gamma_2}{\gamma_1}, r_{13} = \frac{\gamma_1}{\gamma_3}$ and $r_{23} = \frac{\gamma_2}{\gamma_3}$. Hence, $r_{23} = r_{13} \times r_{12}$. If we change r_{13} into $1/r_{13}$ then the ratios are again consistent. The following algorithm does that for all ℓ.

`For` $i = 2 : L$

  `For` $j = 3 : L$

    `If` $\frac{r_{1j}}{r_{1i}} \notin \left\{r_{ij}, \frac{1}{r_{ij}}\right\}$ `then` $r_{1j} \leftarrow \frac{1}{r_{1j}}$
    `If` $\frac{r_{1j}}{r_{1i}} = \frac{1}{r_{ij}}$ `then` $r_{ij} \leftarrow \frac{1}{r_{ij}}$
    `end`

  `end`

`end`

This algorithm guaranties that $r_{\ell\ell'} = \frac{\gamma_\ell}{\gamma_{\ell'}}$ for all ℓ, ℓ' or $r_{\ell\ell'} = \frac{\gamma_{\ell'}}{\gamma_\ell}$ for all ℓ, ℓ'.

Second, assume that $r_{\ell\ell'} = \frac{\gamma_\ell}{\gamma_{\ell'}}$. Then

$$\sum_{\ell'=1}^{L} \frac{1}{r_{\ell\ell'}} = \frac{1}{\gamma_\ell},$$

and $\gamma_\ell = \left(\sum_{\ell'=1}^{L} \frac{1}{r_{\ell\ell'}}\right)^{-1}$. Moreover,

$$\nu_\ell = \left(\sum_{\ell'=1}^{L} \frac{M_{\ell\ell'}}{2M_{\ell\ell}} \frac{\gamma_\ell + \gamma_{\ell'}}{\gamma_{\ell'}}\right)^{-1}.$$

If instead $r_{\ell\ell'} = \frac{\gamma_{\ell'}}{\gamma_\ell}$ then

$$\widetilde{\gamma}_\ell = \left(\sum_{\ell'=1}^{L} \frac{1}{r_{\ell\ell'}}\right)^{-1} \propto \frac{1}{\gamma_\ell}$$

and

$$\widetilde{\nu}_\ell = \left(\sum_{\ell'=1}^{L} \frac{M_{\ell\ell'}}{2M_{\ell\ell}} \frac{\widetilde{\gamma}_\ell + \widetilde{\gamma}_{\ell'}}{\widetilde{\gamma}_{\ell'}}\right)^{-1} \propto \upsilon_\ell \gamma_\ell.$$

Hence, if $\frac{M_{\ell\ell'}}{2M_{\ell\ell}} = \frac{\nu_{\ell'}}{\nu_\ell} \frac{\gamma_{\ell'}}{\gamma_\ell + \gamma_{\ell'}}$, then there is an equivalent parametrization $\frac{M_{\ell\ell'}}{2M_{\ell\ell}} = \frac{\widetilde{\nu}_{\ell'}}{\widetilde{\nu}_\ell} \frac{\widetilde{\gamma}_{\ell'}}{\widetilde{\gamma}_\ell + \widetilde{\gamma}_{\ell'}}$ with $\widetilde{\gamma}_\ell \propto \frac{1}{\gamma_\ell}$ and $\widetilde{\upsilon}_\ell \propto \upsilon_\ell \gamma_\ell$. For each solution $(\lambda_\ell, \nu_\ell, \gamma_\ell)$ there is an observationally equivalent one $(\widetilde{\lambda}_\ell, \widetilde{\nu}_\ell, \widetilde{\gamma}_\ell) \propto \left(\frac{\lambda_\ell}{\gamma_\ell}, \upsilon_\ell \gamma_\ell, \frac{1}{\gamma_\ell}\right)$, where $\propto$ means that second and third components must be normalized so as to sum to one. One additional normalization is required to determine which of the two solutions is the right one. For example, one may prefer ex ante that γ_ℓ be increasing.

D An MM algorithm for the M-step update of transition probabilities

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

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

where $M_{k\ell\neg} \equiv M(\neg|k,\ell)$, $M_{k\ell\ell'} \equiv M(\ell'|k,\ell)$, and

$$\begin{aligned} n_{k\ell\neg}(\beta^{(m)}) &= \sum_i p_i(k|\beta^{(m)}) \, \# \left\{ t : D_{it} = 0, \ell_{it} = \ell, x_{it} = x \right\}, \\ n_{k\ell\ell'}(\beta^{(m)}) &= \sum_i p_i(k|\beta^{(m)}) \, \# \left\{ t : D_{it} = 1, \ell_{it} = \ell, \ell_{i,t+1} = \ell', x_{it} = x \right\}, \end{aligned}$$

where $\#\{\}$ denotes the cardinality of a set and where we reintroduce the control $x_{it} = x$ to remind that we are estimating different parameters for all different control values x.

Parameters $\psi_{k\ell}$ (job finding rate for unemployed) are thus updated as

$$\psi_{k\ell}^{(m+1)} = \frac{n_{k0\ell}(\beta^{(m)})}{n_{k\ell\neg}(\beta^{(m)}) + \sum_{\ell'=1}^{L} n_{k0\ell'}(\beta^{(m)})}.$$

The rest of the 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.[14]

Because the logarithm is concave, we can minorize $M(\neg|k,\ell) \equiv M_{k\ell\neg}$ as follows. With obvious notations, for $\ell = 1, ..., L$,

$$\begin{aligned}
\ln M_{k\ell\neg} &= \ln\left(1 - \delta_{k\ell} - \lambda_k + \lambda_k \sum_{\ell'=1}^{L} \nu_{\ell'}(1 - P_{k\ell\ell'})\right) \\
&\geq \frac{1 - \delta_{k\ell}^{(s)} - \lambda_k^{(s)}}{M_{k\ell\neg}^{(s)}} \ln\left(\frac{1 - \delta_{k\ell} - \lambda_k}{1 - \delta_{k\ell}^{(s)} - \lambda_k^{(s)}} M_{k\ell\neg}^{(s)}\right) \\
&\qquad + \sum_{\ell'=1}^{L} \frac{\lambda_k^{(s)}\nu_{\ell'}^{(s)}(1 - P_{k\ell\ell'}^{(s)})}{M_{k\ell\neg}^{(s)}} \ln\left(\frac{\lambda_k \nu_{\ell'}(1 - P_{k\ell\ell'})}{\lambda_k^{(s)}\nu_{\ell'}^{(s)}(1 - P_{k\ell\ell'}^{(s)})} M_{k\ell\neg}^{(s)}\right).
\end{aligned}$$

Note that both sides of the inequality are equal if $(\lambda_k, \nu_{\ell'}, \gamma_{k\ell}) = (\lambda_k^{(s)}, \nu_{\ell'}^{(s)}, \gamma_{k\ell}^{(s)})$ (no parameter change).

Let

$$\widetilde{n}_{k\ell\ell'}^{(s)} = n_{k\ell\neg} \frac{\lambda_k^{(s)}\nu_{\ell'}^{(s)}(1 - P_{k\ell\ell'}^{(s)})}{M_{k\ell\neg}^{(s)}},$$

where $n_{k\ell\neg}$ is implicitly a function of $\beta^{(m)}$. We omit this reference in the rest of this subsection. This is the predicted fraction of stayers such as home beats visitor ℓ'. Given initial values $\lambda^{(s)}, \nu^{(s)}$ one can update $\gamma^{(s)}$ so as to maximize

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

[14] The MM algorithm works by finding a function that minorizes the objective function and that is more easily maximized. Let $f(\theta)$ be the objective concave function to be maximized. At the m step of the algorithm, the constructed function $g(\theta|\theta_m)$ will be called the minorized version of the objective function at θ_m if

$$g(\theta|\theta_m) \leq f(\theta), \forall\theta, \text{ and } g(\theta_m|\theta_m) = f(\theta_m).$$

Then, maximize $g(\theta|\theta_m)$ instead of $f(\theta)$, and let $\theta_{m+1} = \arg\max_\theta g(\theta|\theta_m)$. The above iterative method guarantees that $f(\theta_m)$ converges to a local optimum or a saddle point as m goes to infinity because

$$f(\theta_{m+1}) \geq g(\theta_{m+1}|\theta_m) \geq g(\theta_m|\theta_m) = f(\theta_m).$$

subject to the normalization $\sum_{\ell=1}^{L}\gamma_{k\ell}=1$.[15] 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

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

That is (taking special care to indices), for $\ell=1,...,L$,

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

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

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

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L}\Bigg(\left(n_{k\ell\neg}\frac{1-\delta_{k\ell}^{(s)}-\lambda_{k}^{(s)}}{\overline{M}_{k\ell}^{(s)}}\right)\ln(1-\delta_{k\ell}-\lambda_k)\\+n_{k\ell0}\ln\delta_{k\ell}+\sum_{\ell'=1}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)}+n_{k\ell\ell'}\right)\ln\lambda_k\Bigg).$$

Let, for $k=1,...,K$,

$$\lambda_k^{(s+1)}=\left[\sum_{\ell=1}^{L}\sum_{\ell'=1}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)}+n_{k\ell\ell'}\right)\right]\left[\sum_{\ell=1}^{L}\left[n_{k\ell0}+n_{k\ell\neg}\frac{1-\delta_{k\ell}^{(s)}-\lambda_{k}^{(s)}}{\overline{M}_{k\ell}^{(s)}}+\sum_{\ell'=1}^{L}\left(\widetilde{n}_{k\ell\ell'}^{(s)}+n_{k\ell\ell'}\right)\right]\right]^{-1},$$

$$\delta_{k\ell}^{(s+1)}=\left(1-\lambda_k^{(s+1)}\right)n_{k\ell0}\left[n_{k\ell\neg}\frac{1-\delta_{k\ell}^{(s)}-\lambda_{k}^{(s)}}{\overline{M}_{k\ell}^{(s)}}+n_{k\ell0}\right]^{-1}.$$

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

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

[15] Notice that for $\ell=\ell'=0$, we have an extra contribution of $(\widetilde{n}_{k00}^{(s)}+n_{k00})\ln\frac{1}{2}$, but it does not matter because it is independent of parameters.

That is

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

For a given value of $\beta^{(m)}$, the sequence $(H(M^{(s)}|\beta^{(m)}))$, driven by the sequence $(\psi^{(m+1)}, \delta^{(s)}, \lambda^{(s)}, \nu^{(s)}, \gamma^{(s)})$, is increasing. The MM algorithm can thus be stopped at any time, not only after convergence, to deliver the updated values of transition parameters, $(\psi^{(m+1)}, \delta^{(m+1)}, \lambda^{(m+1)}, \nu^{(m+1)}, \gamma^{(m+1)})$.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–334.

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

Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics 37*, 3099–3132.

Bagger, J. and R. Lentz (2013). An empirical model of wage dispersion with sorting. *University of Wisconsin–Madison Working Paper*.

Bagger, J., K. L. Soerensen, and R. Vejlin (2013). Wage sorting trends. *Economics Letters 118*(1), 63–67.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016a). Estimating multivariate latent-structure models. *Annals of Statistics 44*, 540–563.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016b). Non-parametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*, 211–229.

Bonhomme, S., T. Lamadon, and E. Manresa (2017). A Distributional Framework for Matched Employer Employee Data. mimeo, University of Chicago.

Card, D., J. Heining, and P. Kline (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics 128*(3), 967–1015.

Celeux, G. and G. Govaert (1992). A classification {EM} algorithm for clustering and two stochastic versions. *Computational Statistics & Data Analysis 14*(3), 315 – 332.

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.

Gassiat, E., A. Cleynen, and S. Robin (2016). Finite state space non parametric hidden Markov models are in general identifiable. *Statistics and Computing*.

Hall, P. and X.-H. Zhou (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics 31*, 201–224.

Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics 5*(1), 123–144.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics 144*, 27–61.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*, 195–216.

Hu, Y. H. and M. Shum (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics 171*, 32–44.

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

Hunter, D. R. and K. Lange (2004). A tutorial on mm algorithms. *Amer. Statist*, 30–37.

Jarosh, G. (2015). Searching for job security and the consequences of job loss. Technical report.

Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica 77*, 135–175.

Lise, J. and J.-M. Robin (2017, April). The macrodynamics of sorting between workers and firms. *American Economic Review 107*(4), 1104–35.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015, May). Firming Up Inequality. NBER Working Papers 21199, National Bureau of Economic Research, Inc.