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

Rasmus Lentz* Suphanit Piyapromdee† Jean-Marc Robin‡

May 8, 2018

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 along with a model of mobility that allows us to quantify sources of sorting. 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. 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 significant sorting on wages and that it is stable over the period. Sorting is established early in careers, increasing during the first decade after which it declines steadily. Counterfactuals demonstrate that sorting is primarily driven by 2 channels: First, a "preference" channel whereby higher wage workers are more likely to accept jobs in higher wage firms. Second, a job finding channel where the job destination distribution out of non-employment is stochastically increasing in the wage type of the worker.

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. Rasmus Lentz gratefully acknowledges financial support from NSF grant SES-1459897.

†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 jeanmarc.robin@sciencespo.fr

1 Introduction

What is the assignment of workers to firms, and how is it realized? In this paper, we estimate a flexible parametric model of wages and mobility with two-sided unobserved heterogeneity using Danish matched employer-employee data. First, this will allow us to measure how wages depend on worker and firm unobserved heterogeneity, and how worker and firm unobserved types correlate across matches in any cross-section (sorting). Second, because we also model the matching process between workers and firms, we will also explain how the estimated cross-sectional match distribution is produced.

Since the seminal contribution of Abowd et al. (1999) (hereafter AKM) the literature that aims to decompose the cross-sectional variance of individual wages into an effect of time-varying observed heterogeneity (essentially tenure and age or potential experience), two separate worker and firm unobserved heterogeneity effects, sometimes a match-specific effect, and a residual component, has focussed on estimating a linear model with fixed effects using Ordinary Least Squares.[1]

The AKM model, in its version without match effects, has been criticized for being too restrictive. Eeckhout and Kircher (2011) claim that non monotonicity in the way wages depend on worker and firm effects may explain the weak correlation that is found when estimating AKM's additive, fixed-effect model. Several recent studies have designed and estimated search-matching models that confirm Eeckhout and Kircher's hypothesis (Lise et al., 2016; Lise and Robin, 2017; Hagedorn et al., 2017; Bagger and Lentz, 2014; de Melo, 2018). In this paper, we do not follow this route as we want to be as agnostic as possible on the way unobserved and firm unobserved heterogeneity determine wages and employment mobility. Our aim is to document relationships so that we can then design appropriate theoretic models.

For the same reason, we also depart from AKM's linearity by following the recent approach of Bonhomme et al. (2017) (BLM), who put no shape restrictions on the way wages and matching depend on worker and firm heterogeneity. This is rendered possible by assuming that worker and firm unobserved types are discrete; that is, workers and firms can be clustered into homogeneous groups of identical workers and firms. The difficulty for the econometrician is that one does not observed which group workers and firms belong to.[2]

Our model and estimation differs from BLM in the following ways. First, consider the probability for a worker of type k and observed time-varying characteristics x who is currently employed at a firm of type ℓ to move to a firm of type ℓ' before the end of the period. The collection all these different probabilities for all k, x, ℓ, ℓ' is soon a very large set. In order to reduce the dimensionality

[1] See Holzer et al., 2004; Martins, 2008; Iranzo et al., 2008; Gruetter and Lalive, 2009; Bagger et al., 2013; Card et al., 2013; Woodcock, 2015; Song et al., 2015; Card et al., 2016. Recent application correct the naive OLS estimator for finite sample biases using the method in Andrews et al. (2008).

[2] As far as we know, BLM and our paper are the only ones which model wages and employment mobility with both worker and firm heterogeneity. There are many papers with only worker heterogeneity (e.g. Altonji et al., 2013) and one paper with only firm heterogeneity (Abowd et al., 2006).

we assume that employed workers meet other employment opportunities at a rate that depends on worker characteristics k and x. The type of the employment opportunity ℓ' 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. So doing, we make the transition probabilities both easier to identify and estimate, and easier to interpret. As in BLM's application, 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.

Second, instead of estimating the model using just two years of data (2002-2004), we use the entire panel of Danish matched employer-employee data for the period 1987-2013. We assume that worker and firm unobserved types do not change over time. However, the way wage and mobility parameters depend on observed and unobserved types is allowed to change over time by periods of three years. For example, the complementarity of worker and firm types in match productivity (wage) can thus change along the business cycle and be subject to structural change. Therefore, a type here conditions wage and mobility in a dynamic way. This allows us to run counterfactual exercises where we remove structural changes to the parameters (time variations). This could also allow, in particular, to calculate ex post earnings present values, like in the recent work of Guvenen et al. (2017), conditional on worker types. The downside of this approach to unobserved heterogeneity is that we may need more types to capture the diversity of possible dynamics.[3]

Third, BLM proceed in two steps for estimation. First, they classify firms using an automatic classification procedure (k-means) based on the distribution of wages within each firm. Then, they estimate the wage and mobility parameters with only worker types unobserved using the EM algorithm. Instead, we propose a Classification EM algorithm in the spirit of Celeux and Govaert (1992), which updates the firm classification after the EM step so as to improve the expected likelihood. So doing, we make use of mobility data, in addition to wages, to classify firm types. This can be important for measuring inequality as two firms offering identical wage but with different levels of job stability would result in different lifetime earnings for workers. Moreover, the mobility parameters involve substantial non-linearity in the likelihood function. The structure of transition probabilities being close to that of a Bradley-Terry model, we formulate an MM algorithm (Hunter, 2004; Hunter and Lange, 2004) that is embedded in the M step and allows a fast solution for mobility parameters - a substantial contribution for the feasibility of the estimator. Monte Carlo

[3]We checked that this modeling assumption was not determining our results by alternatively estimating 9 completely separate models, one for each 3-year subpanel. Both our evaluation of the correlation between worker and firm effects and the variance decomposition barely changed. This by the way also confirms BLM's short-T identification result.

simulations show that this CEM algorithm improves over a one-step k-means.

Nevertheless, we do not view our paper's main contribution as being technical. Our empirical application is rich of interesting lessons. First, we show both in the data and in large-scale simulations (i.e. with a one million workers and 100,000 firms, and using our estimated complex wage and mobility model) that the AKM model tends to underestimate the residual variance and sorting (i.e. the correlation between worker and firm fixed effects, and the contribution of the covariance between worker and firm fixed effects to the log-wage variance), and overestimate the contribution of the worker and firm effects to the overall variance. We find relatively large biases for the OLS estimator of the AKM model that result both from the incidental parameter problem and model misspecification.[4]

Second, our estimates show 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.[5] 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. As a consequence, the cross sectional distribution of match types that results from the first match that is drawn when entering the labor market, and from subsequent employment mobility therefore shows evidence of moderate, yet non negligible 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. This is however a greater correlation than is usually estimated using the AKM model.

Third, our parametric model allows us to run counterfactual simulations that help tracing which parameters are key drivers of sorting. Surprisingly, we find that this correlation level is already 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.

The layout of the paper is as follow. Section 2 describes the model, Section 3 the estimation procedure, Section 4 the parametric specification. Then, Section 5 presents the estimation results, and the last two section presents applications. Section 6 is about measuring sorting and Section 7 about the log-wage variance decomposition.

---

[4] Model misspecification here refers to the fact that, although conditional mean log-wages are not badly approximated by two additive worker and firm effects, there is some evidence of nonlinearity that shows in a non negligible match-specific effect.

[5] For example, we estimate $\mu_{k\ell}$ and $\sigma_{k\ell}$ the conditional mean log-wage and standard deviation for a worker of type k and an employer of type ℓ. The correlation between $\mu_{k\ell}$ and $\sigma_{k\ell}$ across matches (k,ℓ) is estimated around 0.10 for all workers with tenure greater than 100 days.

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.[6]

We assume that employers (firms) can be clustered into L different groups indexed by $\ell \in \{1,\ldots,L\}$ 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: $p(\ell_1|z,k,x_1) = p(\ell_1|k,x_1)$.*

[6]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.

Leaving aside the dependence to 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.

The assumption that z predicts k independently of x_1 is more disputable. First, because some workers in 1987 are drawn from the stock. So x_1 contains tenure and potential experience for these workers. If better workers tend to have higher tenure, for example, then tenure predicts worker type. Second, the association between type and gender and education may vary across cohorts of workers.[7]

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; Gassiat et al., 2016). Recently, Bonhomme et al. (2017) have applied these identification tools to 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 wages and employer's types are generically sufficient to identify the model. In Appendix A we discuss two identification arguments that complement 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.

[7] A more sophisticated assumption would allow k to depend on z and cohort. Then, for experienced workers in 1987, we would translate the current date 1987 and potential experience into the corresponding cohort. Moreover, instead of conditioning on tenure, we would add a term in the likelihood for the probability of initial tenure given worker type, firm type, initial experience and initial date (1987).

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. public/private status) and the unconditional share.

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. 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 firm classification step fail.

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

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

Note that the factors $1/q(\ell|F)$ 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

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

as 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

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

be 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 hence have 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

$$F^{(s+1)} = \arg\max_F \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k|\widehat{\beta}^{(s)}, F^{(s)}) \ln L_i(k; \widehat{\beta}^{(s)}, F). \tag{9}$$

In practice we only search for a firm reclassification that increases the likelihood. To do that we order firms by size from the largest to the smallest. We find $\ell_1^{(s+1)}$ that maximizes equation (9) 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 $\ell_3^{(s)}, \ldots, \ell_J^{(s)}$ 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. Alternatively, we could use the k-means algorithm as in Bonhomme et al. (2017).

4 Empirical specification

In this section, we provide the details of parametric specifications of f and M.

4.1 Wage distribution

Wages are assumed lognormal given match type. Specifically,

$$f(w|k,\ell,x) = \frac{1}{\sigma_{k\ell}(x)}\varphi\left(\frac{w-\mu_{k\ell}(x)}{\sigma_{k\ell}(x)}\right), \tag{10}$$

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 $\sigma^2_{k\ell}$.[8]

The M-step update 5 takes the following form:

$$\mu_{k\ell}^{(m+1)}(x) = \frac{\sum_{i=1}^{I} p_i(k|\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it}=\ell, x_{it}=x\} w_{it}}{\sum_{i=1}^{I} p_i(k|\beta^{(m)}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it}=\ell, x_{it}=x\}},$$

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

(simple averages weighted by posterior type-probabilities).

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 specified as

$$M(\ell'|k,\ell,x) = \lambda_k(x)\nu_{\ell'}(x)P_{k\ell\ell'}(x). \tag{11}$$

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,

$$P_{k\ell\ell'} = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x)+\gamma_{k\ell'}(x)}. \tag{12}$$

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

[8] 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.

same-type job, with little loss of generality, we assume that the worker moves with probability 1/2.[9]

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 given knowledge of unrestricted transition probabilities $M(\ell'|k,\ell) = \lambda_{k\ell}\nu_{k\ell'}P_{k\ell\ell'}$. In this case, 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

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

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 transition probabilities 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.[10]

[9] 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.

[10] 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

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 since graduation), and calendar year.[11] Short tenure is defined as less than 100 weeks of employment, and 26 weeks for non-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.[12] 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 $L = 24$. We indeed obtained greater likelihood gains by categorizing firms more finely than workers.[13] 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

---

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

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

[11] 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.

[12] 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.

[13] We experimented with automatic selection techniques to select K, L such as BIC penalization. For a reason that we do not understand, such likelihood penalization seems to work better for K (workers) than for L (firms). A deeper theoretical analysis of the CEM algorithm (on a simpler specification) is necessary to understand it better.

estimate

$$\mu_{k\ell}(x) = c(x) + a_k + b_\ell + \varepsilon_{k\ell}(x),$$

weighting each cell (k,ℓ,x) by $p(k,\ell,x) = \frac{1}{NT}\sum_{i,t} p_i(k)\mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}$, 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. 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 plotting for example $\mu_{k\ell}(x)$ against $F(k), F(\ell)$ for a given x (year, tenure and experience) over using k and ℓ for the 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 (i.e. $\ell = 10, 11, 13, 14, 16, 20, 21$) but 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 that firm size and vacancy posting are related (as one would expect), the classification will group firms by size, in addition to wage and mobility patterns. Moreover, $F(k)$ is invariant to the labelling of worker types, whether $k = 1,2,3$ or $k = 1,5,7$, both choices deliver the same $F(k)$.

Figure 1: Cross-sectional distributions of types

(a) Marginal distribution of k — (b) Distribution of ℓ across workers — (c) Distribution of ℓ across firms

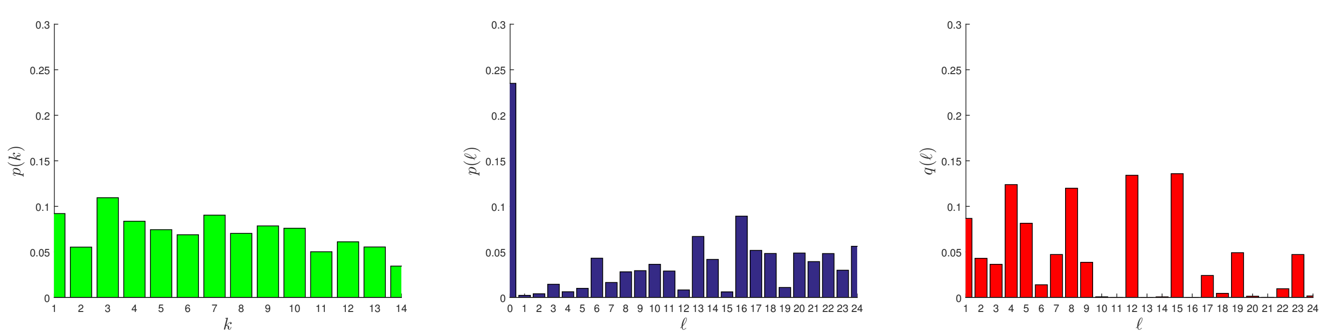


Table 1 shows how observed characteristics z, gender and education, correlate with unobserved worker types (mean posterior probabilities by observed type). There is a strong correlation between types and observed characteristics. High types (large k) tend to be male and highly educated

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

Table 2: Firm Characteristics

Firm type	Sector			Group size	Avg no. spells	%no. spells
ℓ	No cat.	Public	Private			
1	0.14	0.10	0.76	31781	2	0.66%
2	0.07	0.05	0.88	15729	8	1.13%
3	0.09	0.04	0.87	13310	26	3.05%
4	0.16	0.05	0.79	45374	3	1.04%
5	0.07	0.06	0.87	29810	7	1.94%
6	0.15	0.11	0.74	5093	150	6.7%
7	0.03	0.07	0.91	17265	16	2.42%
8	0.36	0.05	0.58	43921	12	4.67%
9	0.08	0.13	0.79	14113	35	4.35%
10	0.05	0.76	0.18	251	1787	3.95%
11	0.18	0.55	0.27	49	7581	3.27%
12	0.04	0.06	0.90	49125	3	1.25%
13	0.67	0.14	0.19	36	23272	7.37%
14	0.39	0.15	0.46	216	3064	5.83%
15	0.20	0.04	0.76	49754	2	0.99%
16	0.17	0.54	0.29	35	30280	9.33%
17	0.06	0.15	0.79	8834	84	6.51%
18	0.06	0.10	0.83	1621	449	6.4%
19	0.03	0.04	0.93	18001	15	2.33%
20	0.07	0.15	0.78	492	1344	5.82%
21	0.11	0.08	0.80	61	7509	4.03%
22	0.10	0.10	0.80	3503	201	6.2%
23	0.16	0.06	0.78	17255	27	4.07%
24	0.07	0.11	0.82	589	1292	6.7%

workers.[14]

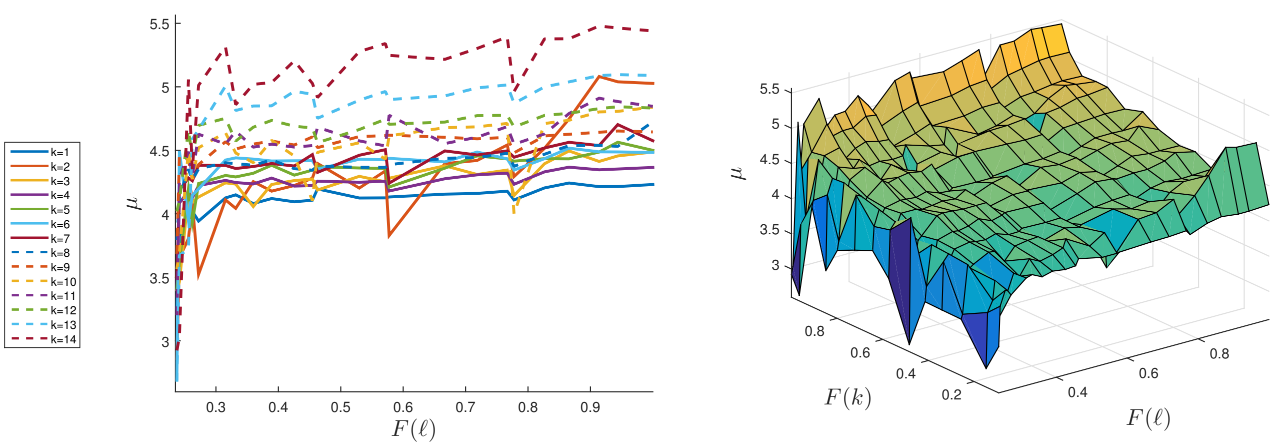
Table 2 repeats the operation for firms. Strikingly, most groups gather many firms that hire very few workers, and a few groups gather a small number of firms that hire many workers. For example, more than half of the 49 firms in group $\ell = 11$ and of the 35 firms of group $\ell = 16$ are public establishments and account for 3.27% and 9.33% of the 11.4 million spells in the panel. The 251 firms in the second largest group, $\ell = 10$, are mostly public and account for 3.95% of all spells. Then, the 61 firms in group $\ell = 21$, the 3,505 firms in group $\ell = 22$ and the 5895 firms in group $\ell = 24$ are mostly private and account for 4.0%, 6.2% and 6.7% 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 16, and groups 21 and 22 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).

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



[14]In reality, types may reflect a combination of heterogeneity in skills, productivity and bargaining power. A more detailed analysis controlling for observed characteristics, such as occupation, and further analyzing sources of heterogeneity in types would be interesting future work.

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. 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.

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 $k=2$; see Figure 3).

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

5.4 Mobility

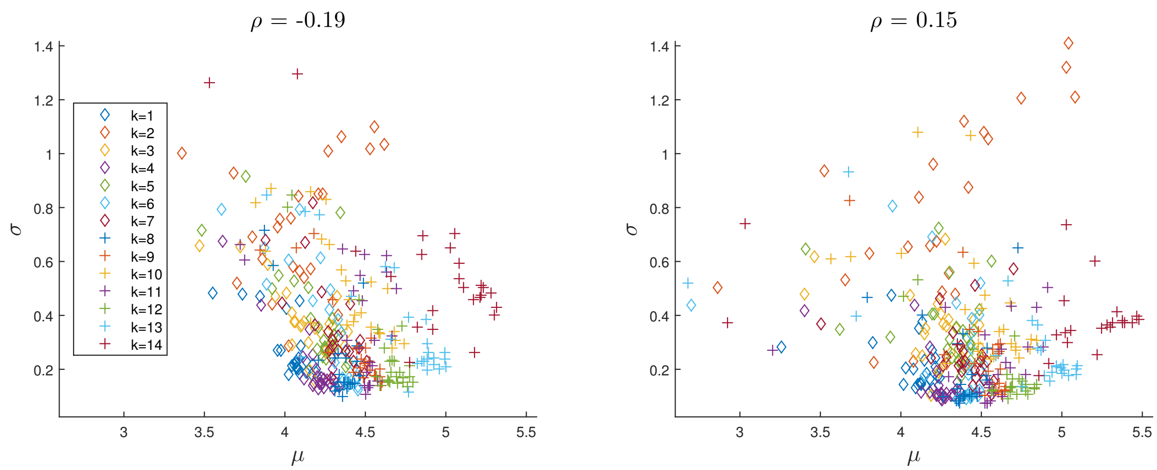
The top row of Figure 4 shows the search intensity parameter λ_k as a function the worker type-rank $F(k)$. Different lines correspond to different levels of experience. The left panel depicts search intensity for low tenure and the right panel displays search intensity for high tenure. Overall, λ_k seems to be increasing in worker type k (with noise), and this relationship more pronounced for younger workers. 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 (2014) 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 for higher tenure workers with little variation across workers of different experience levels. However, this relationship is relatively weak for short tenured workers (see also the correlations in Table 3).

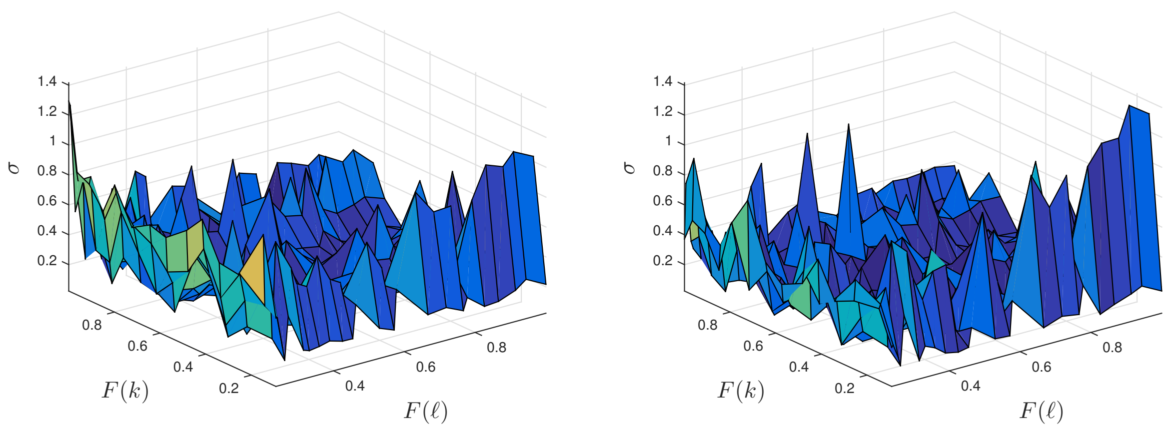
Figure 3: Idiosyncratic dispersion $\sigma_{k\ell}$ (10-15 years of experience)

tenure < 100 weeks   tenure > 100 weeks

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



(b) By worker and firm type-rank



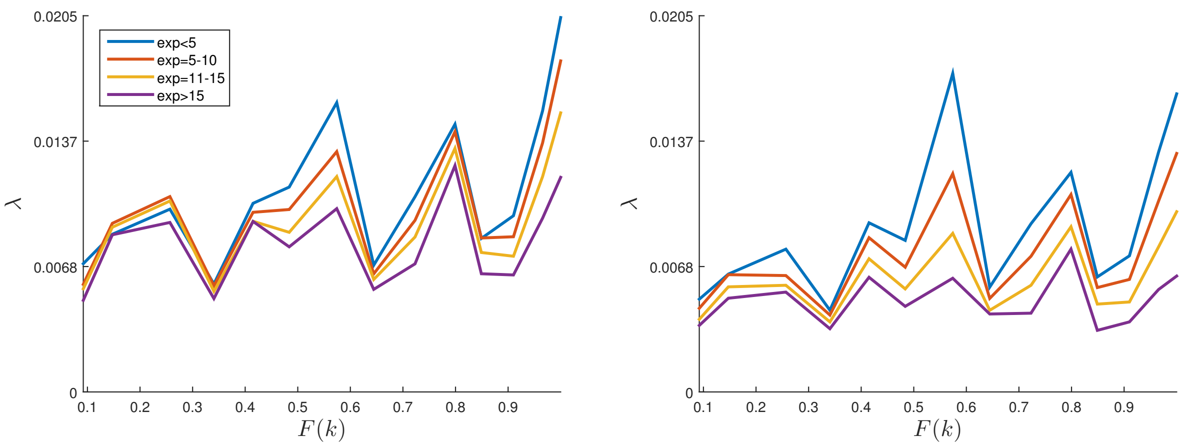
Figures 5 and 6 depict the other mobility parameters $\delta_{k\ell}, \psi_{k\ell}, \gamma_{k\ell}$ by mean wage $\mu_{k\ell}$ and by worker and firm types. We only show the plots for workers with 10-15 years of experience as the different levels of experience do not change the general patterns. The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

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 (2014) 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 mean wage and the preference for the job (or match value), $\gamma_{k\ell}$, is slightly negative

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

tenure < 100 weeks   tenure > 100 weeks

(a) λ_k: probability of meeting outside firm



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

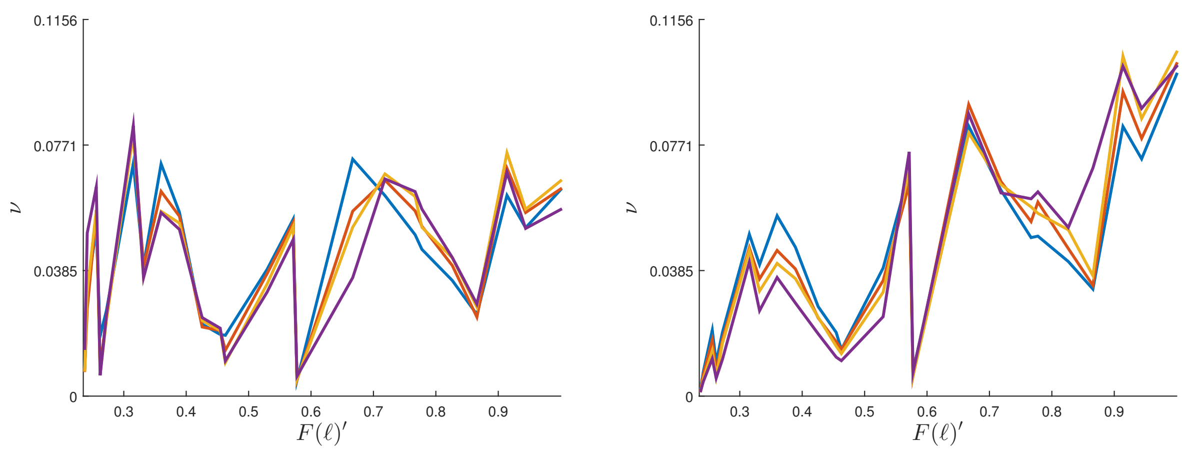


Figure 5: Mobility parameters by mean wage $\mu_{k\ell}$ (10-15 years of experience)

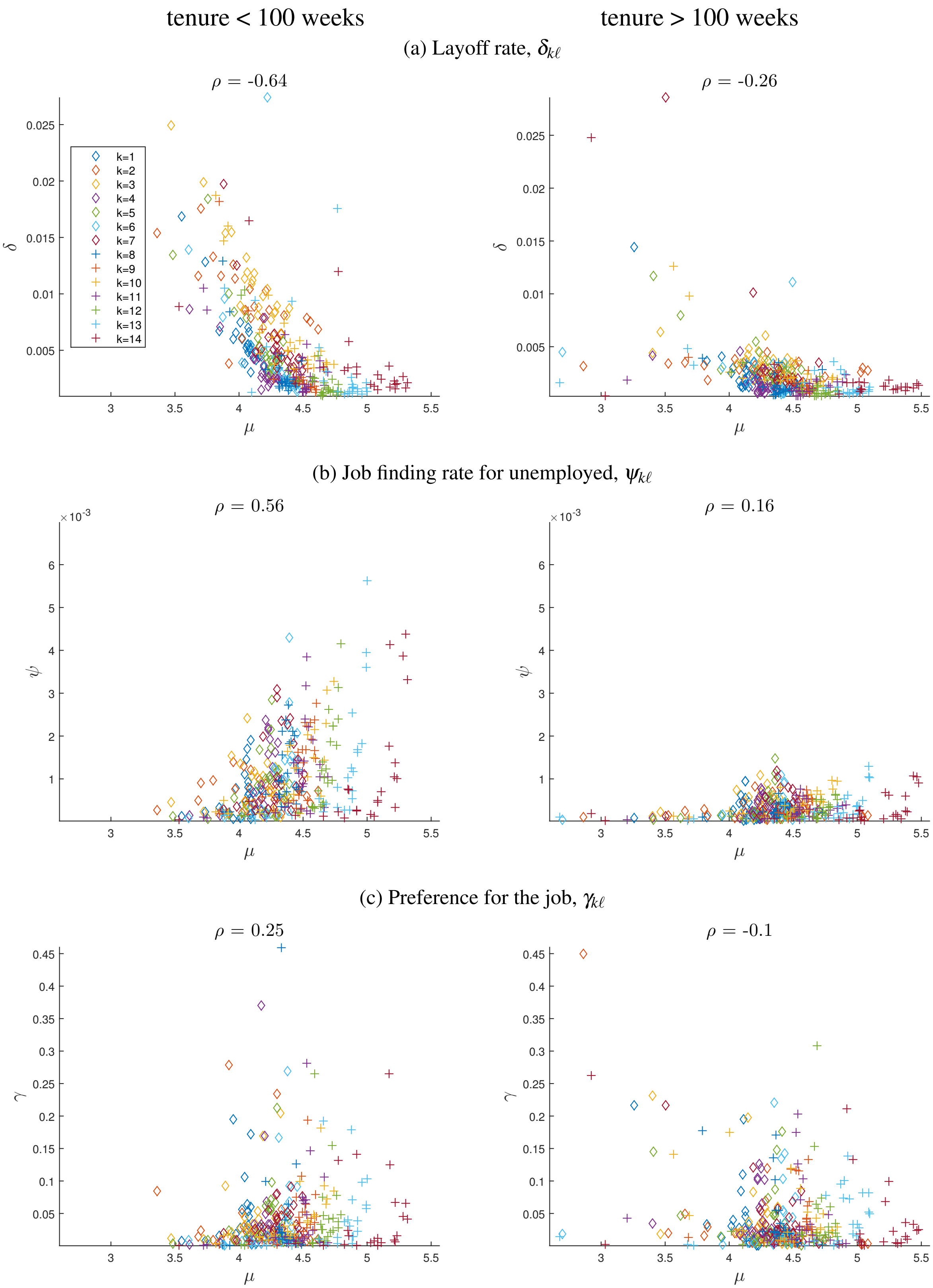
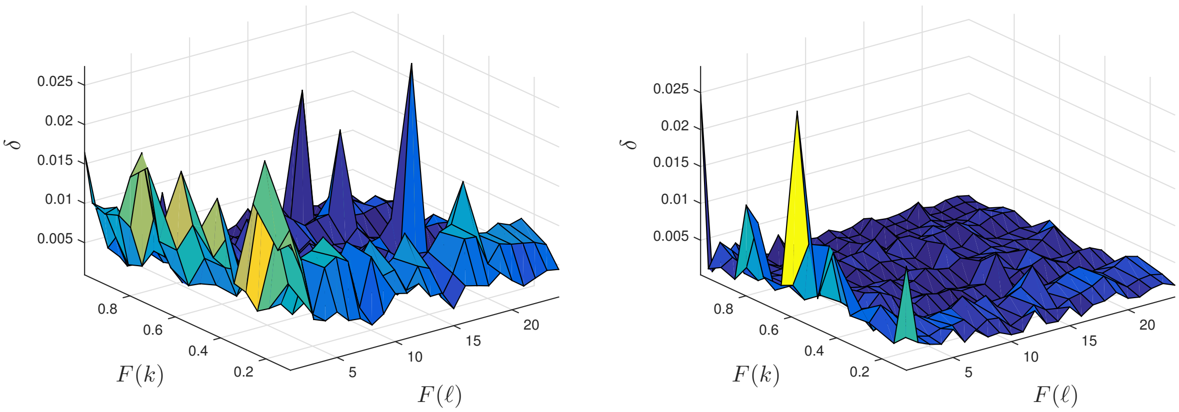


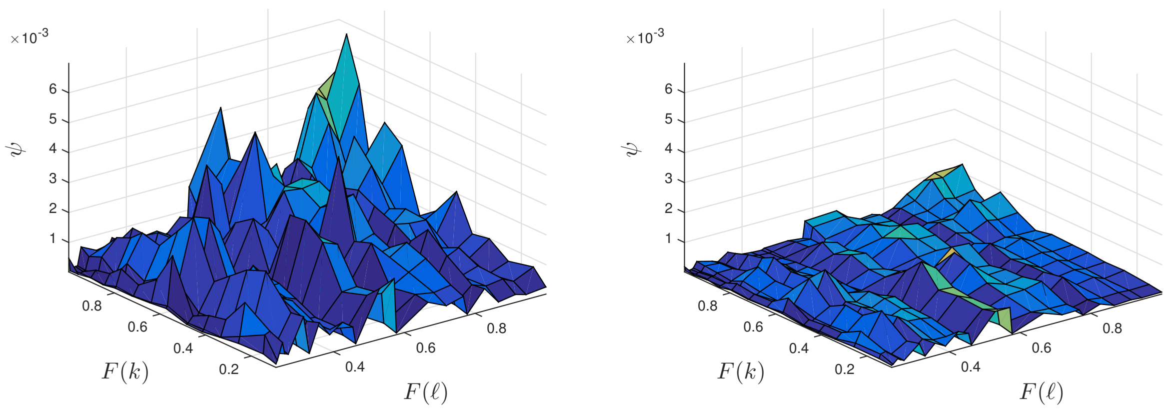
Figure 6: Mobility parameters by worker and firm type-rank (10-15 years of experience)

tenure $<$ 100 weeks tenure $>$ 100 weeks

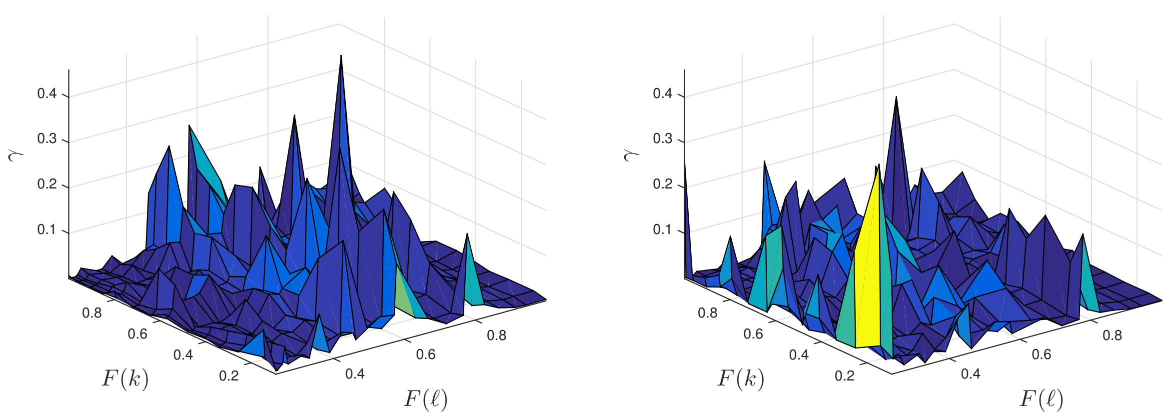
(a) Layoff rate, $\delta_{k\ell}$



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



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



for 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, particularly for long tenure workers. 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 $\lambda_k, \nu_\ell, \gamma_{k\ell}$ 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 rather than the current firm type.

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 (2014). 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, as we have just seen. The aim of this section is first to provide a measurement of sorting, and second to understand how it is built.

6.1 Distributions of matches

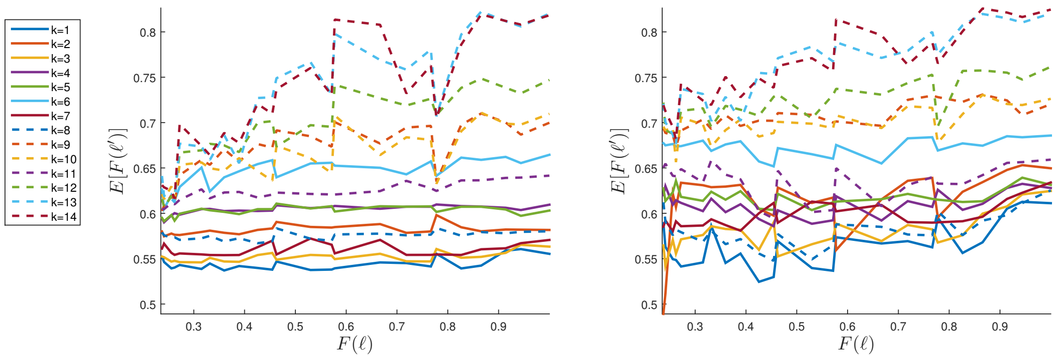
To illustrate the estimated sorting, Figure 8 shows the joint distribution of workers and firms relative to the matching probability under the assumption of independence, $c[F(k), F(\ell)] \equiv \frac{p(k,\ell)}{p(k)p(\ell)}$. This is a copula density. There is evidence of positive assortative matching. 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 calculate the correlation coefficient between the cardinal measures of worker and firm types we obtained in the wage projection. The worker type measure is $a_k(x)$ and the firm type measure is

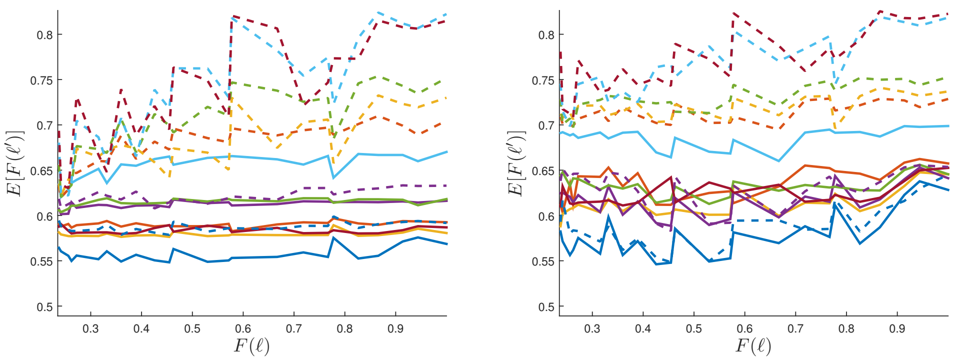
Figure 7: Conditional mobility across firm types

tenure < 100 weeks   tenure > 100 weeks

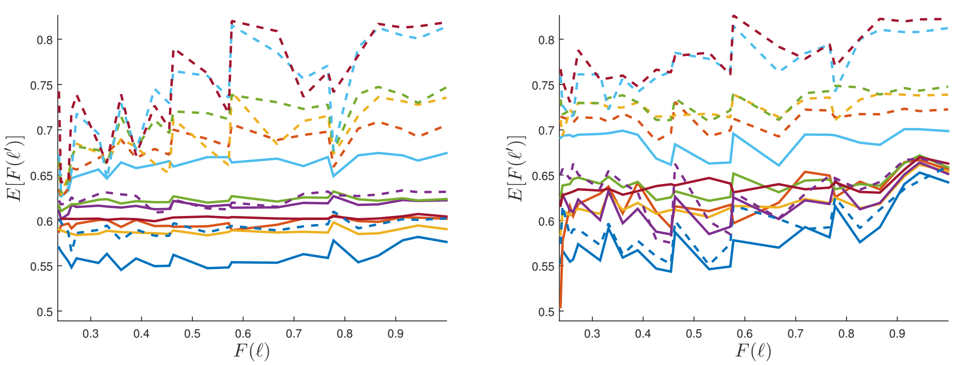
(a) Less than 5 years of experience



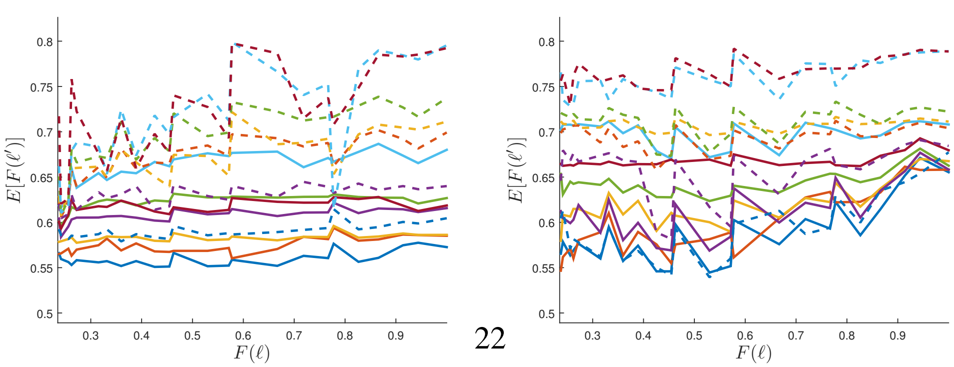
(b) 5-10 years of experience



(c) 10-15 years of experience



(d) at least 15 years of experience



$b_\ell(x)$, obtained by projecting $\mu_{k\ell}(x)$ on k and ℓ separately for each x, each case (k,ℓ) being weighted by $p(k,\ell|x) \propto \sum_{i,t} p_i(k)\mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}$. 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. 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. Also, sorting seems a bit stronger for younger workers, which is consistent with the search ladder being steeper for younger workers.

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, with the usual projection on x, k, ℓ, i.e. without interacting k and ℓ with tenure and experience, we find a fairly modest increase in the correlation coefficient over time from 0.25 to 0.3 (see Figure 10, line "Observed"). 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.

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 $M_{k\ell\ell'}(x)$;[15] 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). It is somewhat surprising to find that mobility is not a mechanism that largely increases sorting, instead mobility just helps maintain sorting close to its initial level.

Synthetic cohorts

To illustrate the interaction between experience and tenure on mobility patterns, we construct a synthetic cohort for each of our 9 time periods. The cohort is initialized with zero experience and tenure. Its initial allocation is drawn according to the estimated $m(\ell|k,x)$ where the x category is the one corresponding to zero experience, zero tenure and the time period in question. The cohort

[15] From $M_{k\ell\ell'}(x)$ calculate $m^\infty(\ell|k,x)$ as the ergodic distribution associated with the Markov chain where we also allow tenure status to reset. 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 8: Copula $c[F(k),F(\ell)] \equiv \frac{p(k,\ell)}{p(k)p(\ell)}$

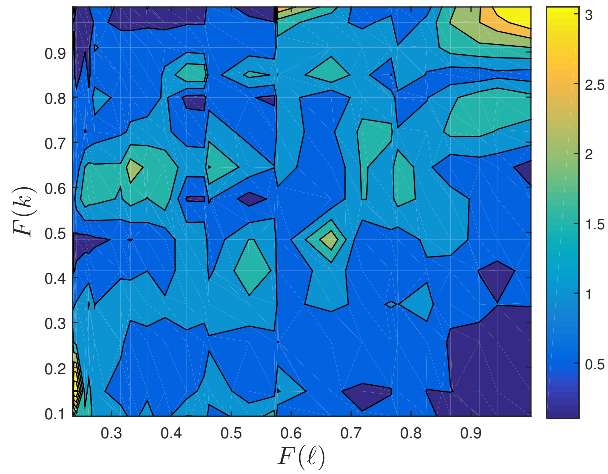
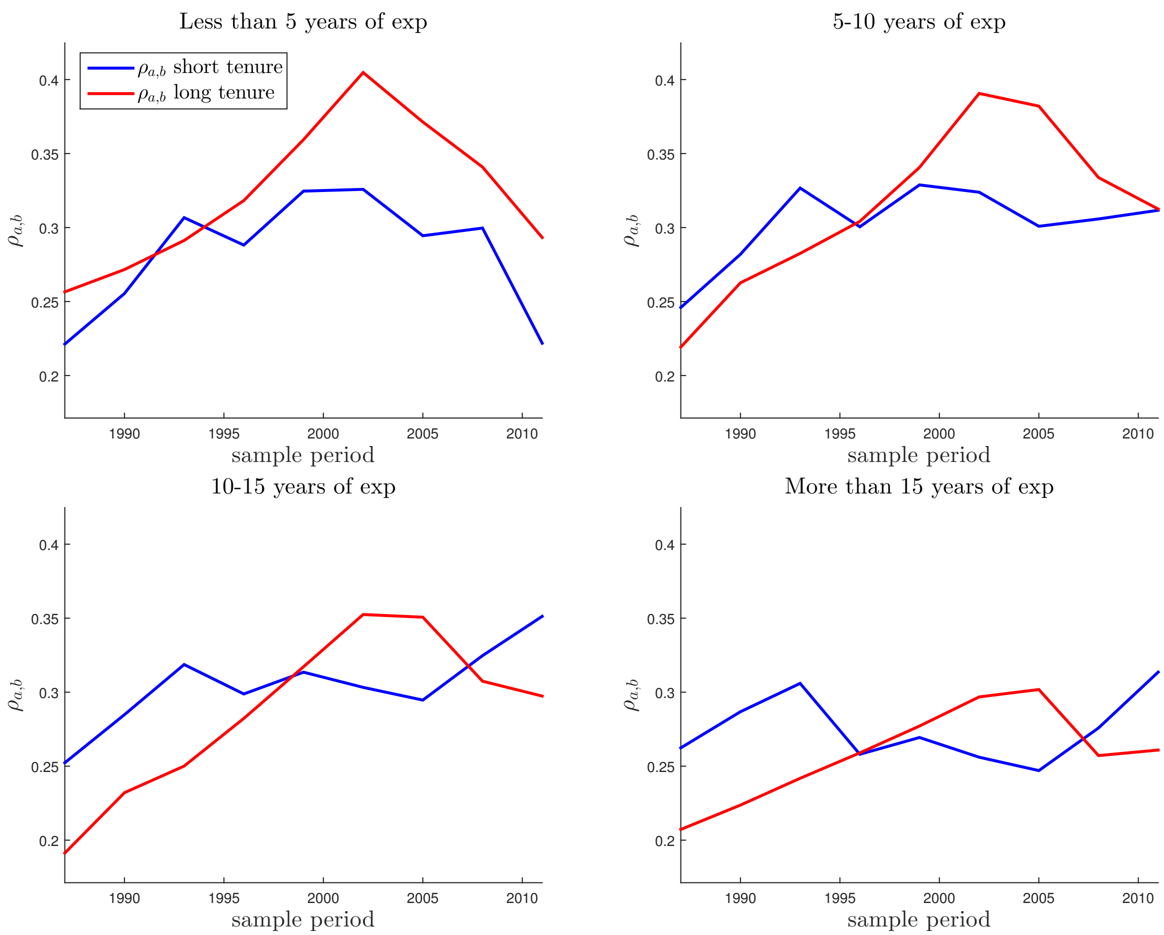


Figure 9: Sorting by tenure and experience



is then simulated forward 30 years with experience and tenure evolving endogenously. Time is held counter factually constant to avoid conflation of experience effects with time effects.

Figure 11 shows the correlation between worker and firm wage effects among employed workers for the 9 different cohorts by experience. Already at the outset, the cohorts are significantly sorted. It is then a general pattern that during the first 10 years of a cohort's life, mobility patterns strengthen sorting, and the cohort's maximal sorting level is attained already after about 10 years. Sorting subsequently weakens as the cohort ages further.

Figure 12 further illustrates the sorting patterns, in this case for the 1990-92 cohort. The left hand shows the average firm location, measured by the firm's wage effect, by experience and worker type. The cohort is positively sorted and it is seen that higher worker types are on average matched with higher wage firms. Furthermore, the increased sorting over the first 10 years of the life cycle of the cohort seems to be primarily driven by better workers improving their firm location.

The right hand graph in Figure 12 shows the considerable sorting on the extensive margin that is also implied by the estimated model. It is seen that low wage workers have significantly higher non-employment rates early in life than high wage workers. All worker types increase their employment rates with age, but it is particularly pronounced for the low types resulting in much less employmnet rate dispersion across worker types for high experience workers.

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. Each counterfactual changes a particular mobility variable by removing one dimension of variation (such as worker-type, firm-type, tenure, experience and time variation) and setting the variable to its weighted average in a given dimension while holding the others parameters constant. We simulate 1,000,000,000 workers to calculate the counterfactual steady state match probability $p^{\infty}_{alt}(k,\ell)$. We then use $p^{\infty}_{alt}(k,\ell)$ to re-calculate the correlations between the benchmark cardinal measures of worker a_k and firm b_ℓ types, obtained from the original steady-state distribution.[16] Table 4 displays the percentage change of counterfactual correlations, relative to the benchmark case, for both unconditional and conditional on experience.

In Figures 13 and 14, we only display the results of the unconditional correlations since in most exercises, the counterfactual correlations across age group are qualitatively similar (however with an exception of one exercise of which will be discussed shortly). Figure 13a 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.

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

Table 4: Average counterfactual sorting over time

Benchmark	$\rho_{a,b}$				
	Average	<5 exp	5-10 exp	10-15 exp	15> exp
	0.28	0.30	0.31	0.26	0.27

	% change				
Counterfactual	Average	<5 exp	5-10 exp	10-15 exp	15> exp
No k variation in $\gamma_{k\ell}$	-63.47	-61.05	-63.57	-66.50	-63.90
No k,ℓ variation in $\gamma_{k\ell}$	-69.89	-70.37	-69.55	-64.72	-70.05
No k variation in $\psi_{k\ell}$	-33.63	-29.70	-32.89	-33.41	-37.90
No ℓ variation in $\psi_{k\ell}$	-42.30	-28.26	-36.92	-49.77	-46.27
No ℓ variation in ν_ℓ	-57.59	-33.16	-49.50	-74.68	-65.30
No k variation in λ_k	-2.87	-3.69	-3.06	-2.81	-1.92
No E-E transition ($\lambda_k = 0$)	-39.49	-41.72	-42.22	-40.60	-27.92
No k variation in $\delta_{k\ell}$	0.37	-4.85	-3.43	3.56	2.61
No ℓ variation in $\delta_{k\ell}$	-2.18	1.08	-2.09	1.92	-4.87
No k variations in $\delta_{k\ell}$ given no k variation in $\psi_{k\ell}$	-38.06	-42.35	-41.50	-35.36	-37.33
No k variation in $\gamma_{k\ell}$ given no ℓ variation in ν_ℓ	-102.1	-95.61	-97.43	-116.99	-100.13
No tenure effect in $\gamma_{k\ell}$	-6.85	-3.21	1.16	-8.20	-11.40
No tenure effect in λ_k	-0.50	-0.16	0.03	-2.37	-0.69
No tenure effect in ν_ℓ	-2.35	-0.71	-1.49	-0.55	-4.31
No experience effect in $\gamma_{k\ell}$	-2.76	-13.83	-11.20	3.90	7.45
No experience effect in λ_k	-2.26	-2.13	-1.60	-2.81	-2.89
No experience effect in ν_ℓ	-0.56	1.23	-0.01	-2.58	-1.16
No E-E transition ($\lambda_k = 0$)	-39.49	-41.72	-42.22	-40.60	-27.92
No E-U transition ($\delta_{k\ell} = 0$)	-10.52	23.83	0.34	-22.12	-13.29

Figure 13b 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, $\gamma_{k\ell} = \overline{\gamma}$. There is no substantial difference between the two counterfactuals, maybe because in both cases there is no specific (k, ℓ) complementarity in transition probabilities.

Turning to the other significant source of sorting in the estimated model, Figure 13c 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 13d 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. Similarly, when we set λ_k to zero such that workers keep their first jobs they draw out of unemployment until the next layoff, sorting is reduced by about 40 percent (see Figure 13e).

Figure 13f shows that the removal of heterogeneity in the sampling probability of firm types $v_{\ell'}$ 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 14a 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 14b&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. Therefore, differential layoff differences across worker types would not affect sorting either. Figure 14d 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.

Finally, we assess the extent to which the transition from employment to unemployment impacts sorting by setting $\delta_{k\ell}$ to zero. Figure 15 displays the conditional correlations by age group. As can be seen in Figure 15a, when workers only move from job to job, sorting rises among young workers, while sorting reduces for workers in other age groups. This is consistent with our finding that sorting is largely built early on in one's career and become less strong as workers age.

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 16, 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 match wage variance, $\mathrm{E}\sigma^2$, and between match wage variance, $\mathrm{V}\mu$, respectively. As can be seen, the between match wage variance is rising proportionately over time so that by the end of the sample, in the early 2010s, the within match wage variance accounts for 62.1% of overall wage variance, from 47.6% at the end of the 1980s. The residual variance, the average of the idiosyncratic component $\sigma_{k\ell}(x)$, conversely decreased from around 52.4% to 37.9%.

Note that we estimate considerably more idiosyncratic wage dispersion than is usually the case using the AKM model (Table 5). 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 29% between 1987 and 2007, and 35% after 2007), but its importance as a fraction of between match variation is decreasing over time (from 61.2% to 58.5%). 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 (Card et al., 2013; Song et al., 2015), although in a much more subdued way. The firm effect has remained third in contribution after the worker effect and what we conventionally call "sorting" until the mid-2000, explaining about 5% of the log wage variance. We observe a big drop after 2005 benefiting to the match-specific effect $V(\widetilde{\mu})$ and the effect of age and tenure $V(\overline{\mu})$. It is difficult to say if this corresponds to a structural change or if it is just some anecdotic phenomenon. Finally, there is evidence that worker and firm heterogeneity impact mean wages in a non additive way. The match-specific effect $V(\widetilde{\mu})$ has doubled in size between 1995 and 2011. Its contribution to log wage variance is now comparable to that of the worker effect.

Summing up, 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, sorting and residual effects are different from the estimates in Bonhomme et al. (2017) using Swedish matched employer-employee data. They estimate a greater correlation between worker and firm types, less residual variance, a much greater contribution of the person effect and a lower contribution of the firm effect. However, note that they estimate their more flexible model (as far as mobility is concerned) on just two years of data, whereas we use 26 years of data to identify worker and firm heterogeneity.

Table 5: Log-wage variance decompositions (percents)

		AKM											BLM	Our estimator			
		FR	US1	AU	US2	IT	BR	DE		US3		DK	SW	Denmark			
		76-87	84-93	90-97	90-99	81-97	95-05	85-91	02-09	80-86	07-13	87-13	02-04	87-89	99-01	11-13	average
Residual	$\mathrm{E}\sigma^2$	15.8	9.3	5.0	9.3	15.1	7.0	7.7	5.0	20.3	14.9	39.0	25.2	52.4	52.6	37.9	50.5
Person effect	$\mathrm{V}a$	76.9	81.6	66.3	63.7	43.9	60.0	61.0	51.2	47.5	52.8	42.9	60.1	29.1	28.4	36.3	30.2
Firm effect	$\mathrm{V}b$	30.2	19.2	37.0	15.4	13.1	26.9	18.5	21.2	16.0	11.9	11.6	2.5	5.3	5.3	3.8	5.0
Cross effect	$2\mathrm{Cov}(a,b)$	-27.2	-2.0	-22.4	0.62	2.1	3.2	2.3	16.4	1.6	7.1	3.3	12.2	5.6	7.3	7.3	6.9
Match effect	$\mathrm{V}\widetilde{\mu}$				5.1			2.6	2.3					5.2	3.4	7.5	4.6
Observed heterogeneity	$V\overline{\mu}$	6.8	52.0	3.1	4.0	7.5	3.0	10.7	2.8	7.6	7.2	1.8		2.7	2.7	6.3	3.4
	$2\mathrm{Cov}(a,x)$	-3.1	-69.0	9.4	0.64	15.5		-2.6	0.70			0.88		-0.21	-0.12	0.14	-0.10
	$2\mathrm{Cov}(b,x)$	0.7	9.0	1.7	1.25	2.6		1.6	1.7			0.52		-0.03	0.40	0.71	0.41
Sorting	$\mathrm{Corr}(a,b)$	-28.3	-2.5	-22.7	1.0	4.4	4.0	3.4	24.9	2.9	14.2	7.4	49.1	22.5	30.0	31.0	28.9

Notes: All decompositions under the label AKM refer to estimations of the AKM model with or without a match-specific component on various matched employer-employee data using the method of Abowd et al. (2002). Columns FR (France) and US1 come from Table 2 in Abowd et al. (2002). Column AU (Austria) comes from Table 4 in Gruetter and Lalive (2009). Column US2 results from Tables 5 and 6 in Woodcock (2015). They use the LEHD data discussed in Abowd et al. (2009). Column IT (Italy) refers to Table 4 in Iranzo et al. (2008). Column BR (Brazil) refers to Table 1 in de Melo (2018). Columns DE (Germany) are calculated from Table III in Card et al. (2013). Columns US3 result from our own calculations from Table 2 in Song et al. (2015). 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". Column DK, under the AKM label, corresponds to our own estimation of the AKM model on entire 10987-2013 sample. Column BLM refers to Bonhomme et al. (2017)'s estimation for Sweden, 2002-04 (Table 2).

Notice finally that the share of log wage variance explained by the correlation between worker and firm effects, and this correlation give two different measurements of “sorting”. The latter one being relative to the variances of worker and firm effects, we obtain a correlation that is three or four times larger than with the AKM model. Recently, Borovičkovà and Shimer (2017) have proposed a different way of calculating this correlation based on mean wage per worker and per firm. Using their estimator, we find a correlation of 47%, much larger than ours. Understanding the sources of discrepancy is beyond the aim of this paper, but it is certain that this estimator measures something different from the correlation between worker and firm linear effects.

Monte Carlo simulations

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 simulations involve 1,000,000 workers and 100,000 firms where $K = 14, L = 24$. The length of the panel is 7 years. We use the parameters estimated for a given time period and age class, and we simulation worker trajectories, allowing for tenure to change. We hold age fixed in the simulations so as not to worry about workers' entry into and exit from the labor market. We draw initial conditions from the steady-state distribution.

The results can be found in Table 6. Each horizontal section of the table separated by an horizontal line corresponds to a different simulation with a different time period and a different experience group. Within each horizontal panel, the row labeled “true” displays the actual variance decomposition and correlation between worker and firm effects for employed workers. The row “LPR” shows these variance decomposition and correlation after first estimating the parameters using our model and estimation procedure. It is a test of the capacity of our CEM algorithm to recover the true parameters when the model is well specified. As can be seen, our algorithm works very well. The row labeled “AKM” shows the variance decomposition and correlation that are obtained by estimating an AKM model. 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. Lastly, the row labeled “BS” shows Borovičkovà and Shimer (2017)'s calculation of the degree of sorting. The simulations confirm the tendency of the correlation between mean wage per worker and mean wage per firm (with all the adjustments to the naive formula proposed by BS) to overestimate the correlation between worker and firm linear effects. Obviously, this statistic captures more than just the correlation between linear projections of log wages on worker and firm dummies. The role of the match-specific effect, for example, must be better understood.

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

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

We have

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

and

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

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: $2\mathrm{Cov}\left(\Delta\overline{\mu}(x), \Delta a_k\right)$, $2\mathrm{Cov}\left(\Delta\overline{\mu}(x), \Delta b_{\ell\ell'}\right)$.

Results

The results are displayed in Figure 17. 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.

What all this means is that the more precise description of wage and mobility that BLM's approach facilitates does not fundamentally upset the estimation of sorting that the AKM model delivers. A 25% correlation is certainly not large. BLM estimate a larger correlation on Swedish data but they only use two years of data, the bare minimum for identification. We could understand such a small correlation for young workers, but most search-matching models would predict that a stronger correlation should build after a while. Our estimates of the preference for a job (determining which of the current job and an alternative one a worker selects) are very weekly correlated with mean wages (across match types), in particular for older workers. This implies that workers tend to change employment for reasons that relate very little to productivity. Amenities, following one's spouse, skill obsolescence may explain mobility at older age better. More work is needed to understand what is determining employment mobility and how it relates to wages.

Figure 10: Observed, initial and equilibrium sorting

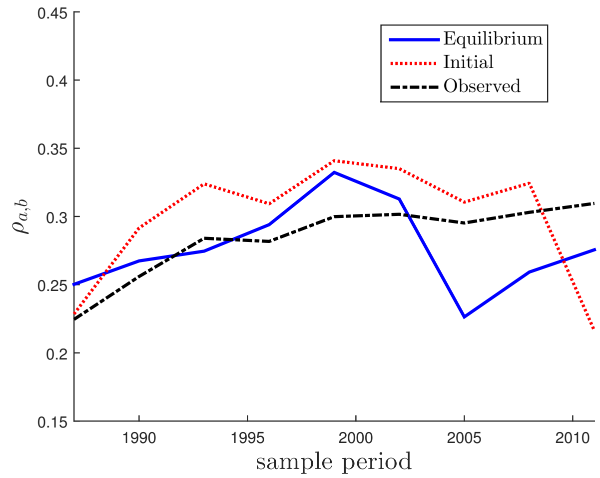


Figure 11: Sorting over the life cycle

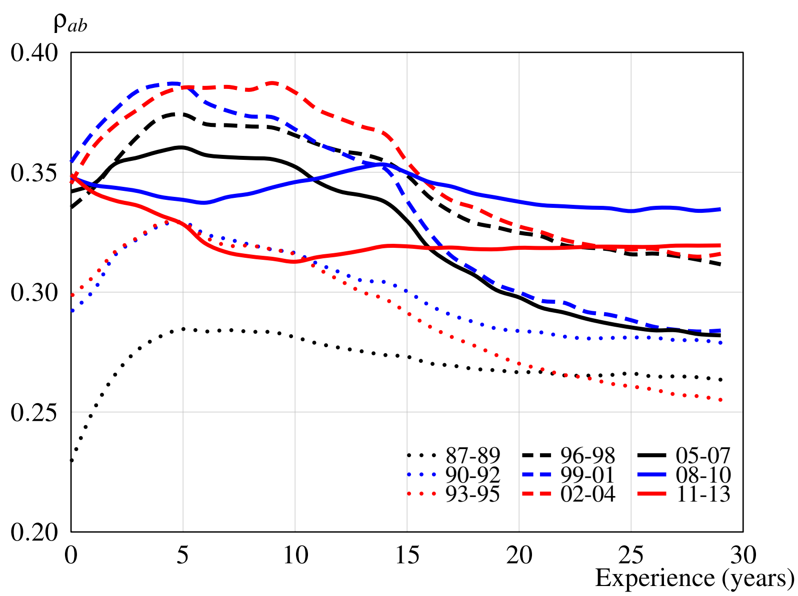


Figure 12: Non-employment and average firm location, 1990-92 cohort.

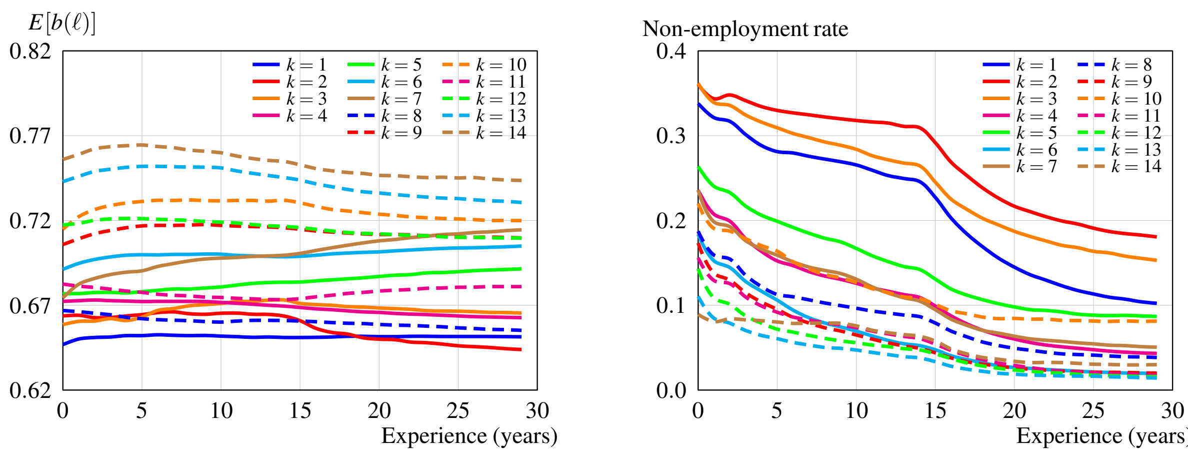


Figure 13: Potent sources of sorting

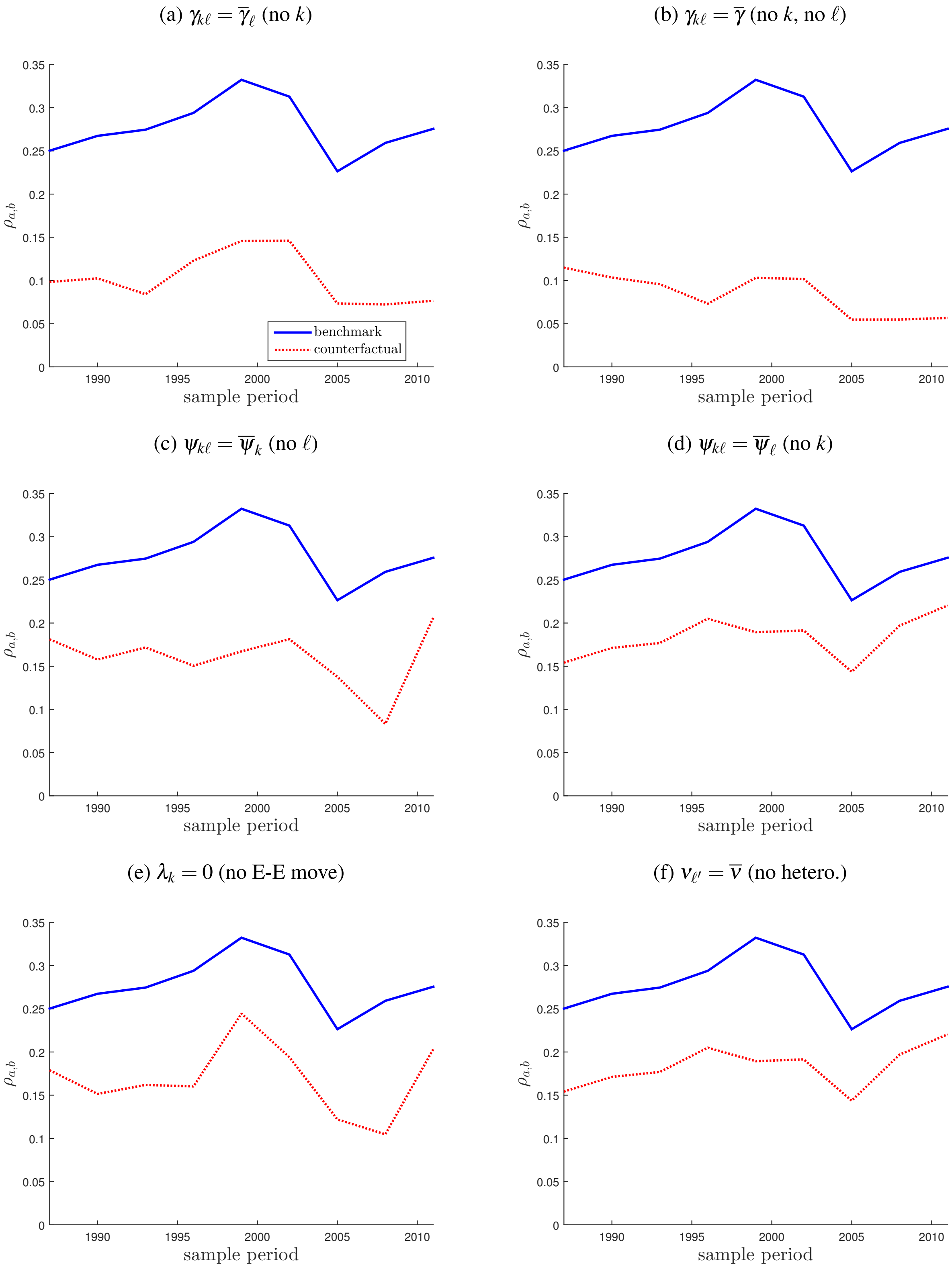


Figure 14: Ineffective sources of sorting

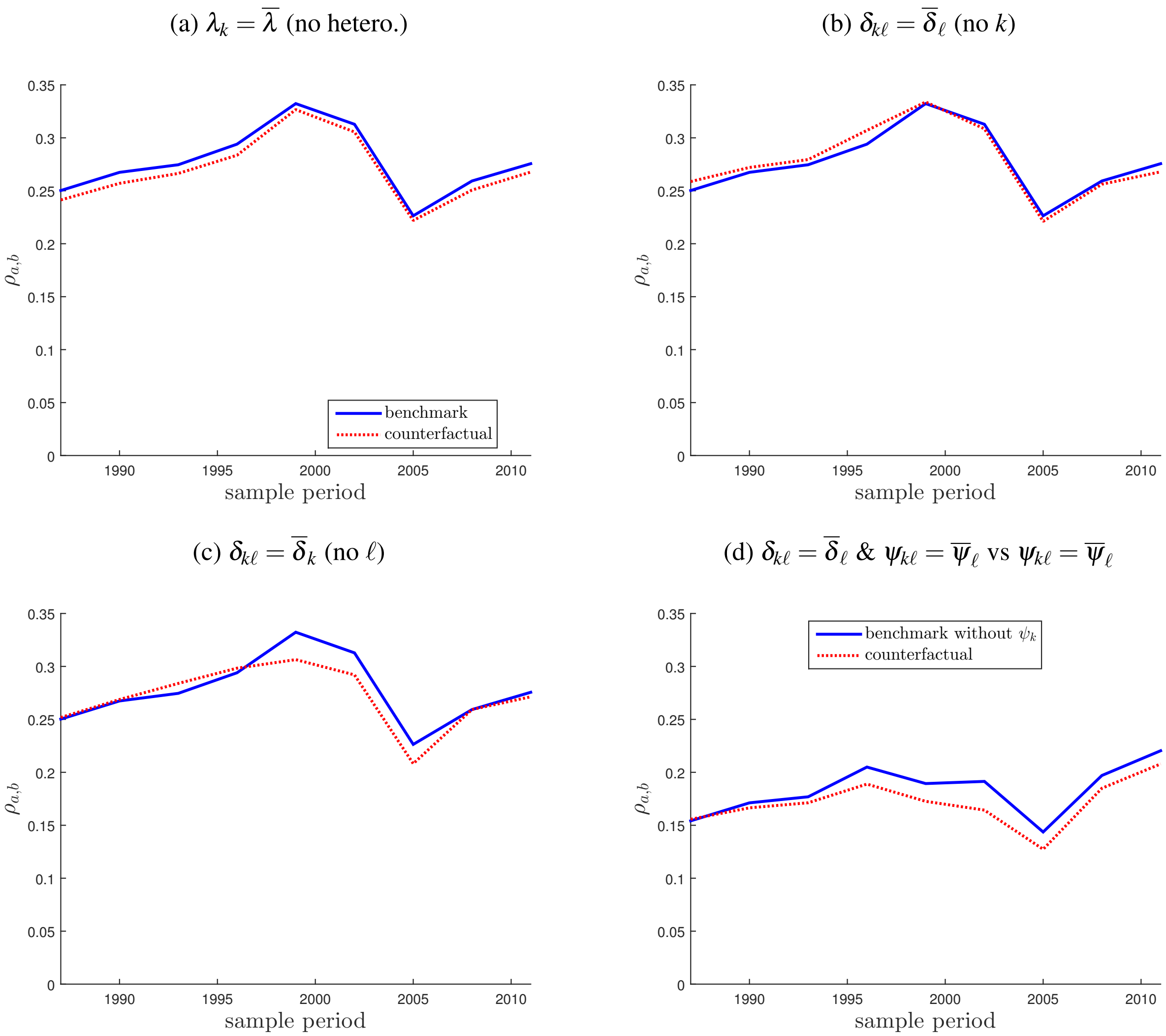
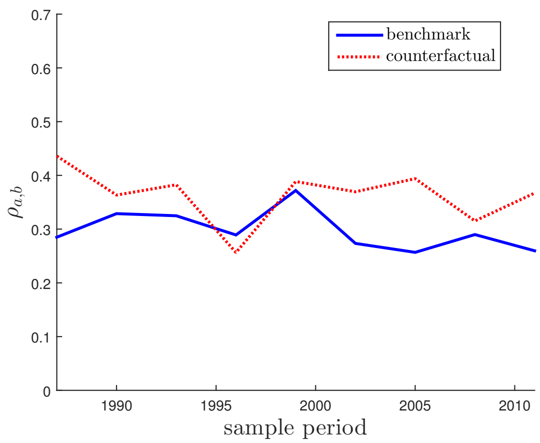
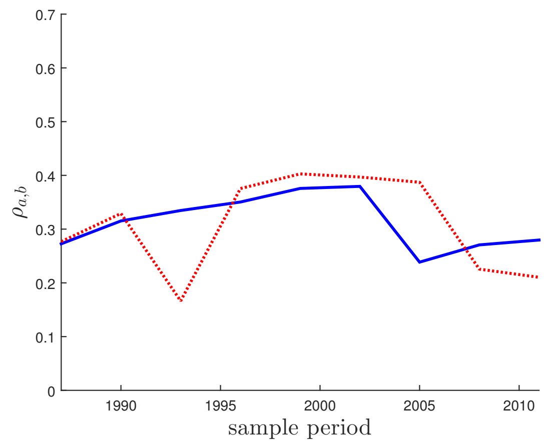


Figure 15: No E-U transition $\delta_{k\ell} = 0$

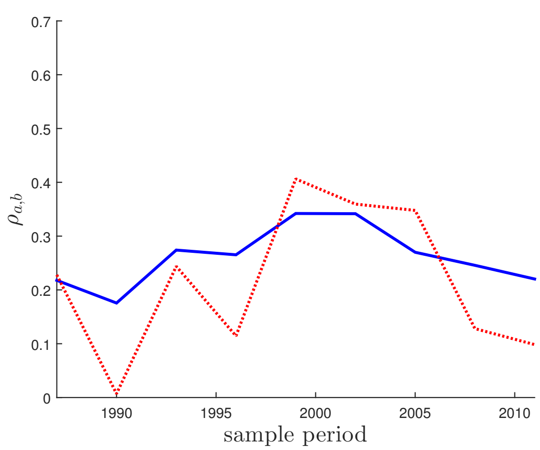
(a) Less than 5 years of exp



(b) 5-10 years of exp



(c) 10-15 years of exp



(d) More than 15 years of exp

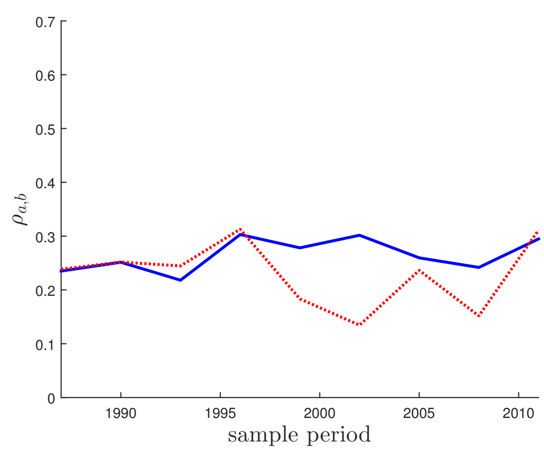
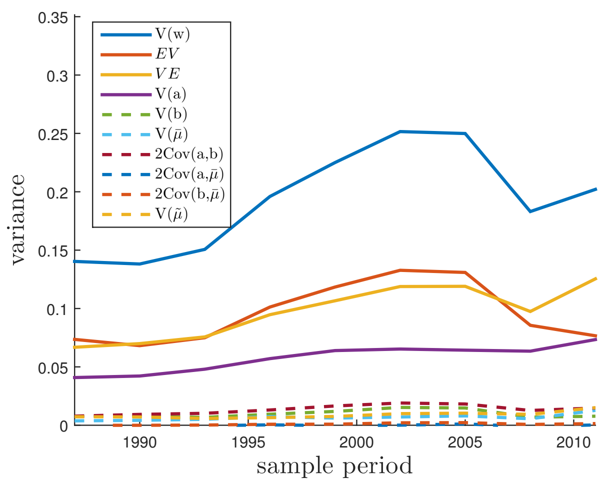
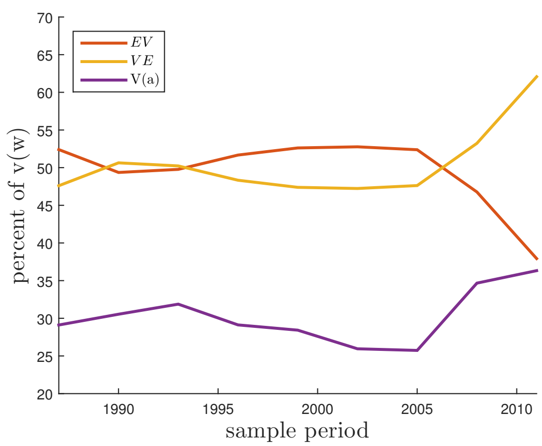


Figure 16: Variance decomposition

(a) Levels



(b) Main components



(c) Firm effect and sorting

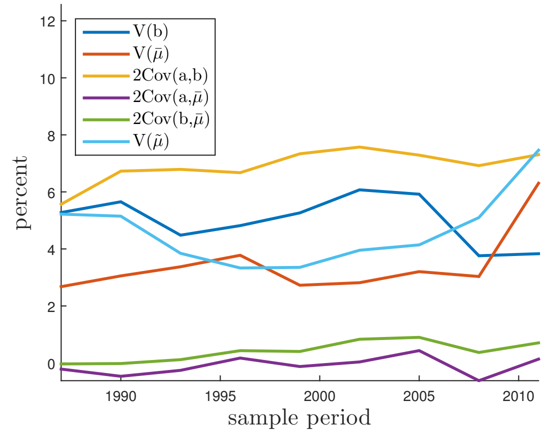
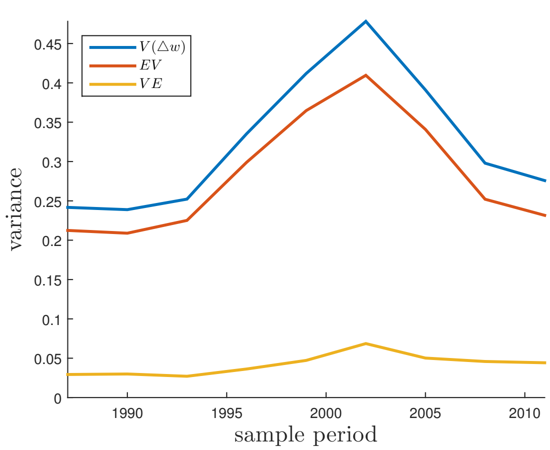


Figure 17: Variance decomposition of wage growth



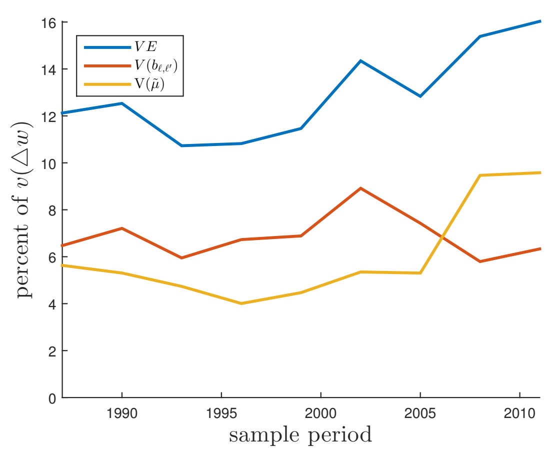


Table 6: Simulated log-wage variance decomposition

Date	Experience	Model	Residual variance	Explained variance	Components of explained variance: Person effect	Firm effect	Sorting	Match effect	Obs. hetero. var	cov		Correlation btw pe and fe
1987-89	<5 yrs	true	67.9	32.1	16.5	5.6	5.1	3.3	1.2	0.2	0.1	26.6
		LPR	68.0	32.0	17.4	6.0	4.5	2.5	1.4	-0.1	0.2	22.2
		AKM	53.2	46.8	35.2	10.5	-0.6		1.3	0.0	0.2	-1.4
		BS										33.8
	5-10 yrs	true	59.2	40.8	25.0	6.2	6.7	2.6	0.3	0.1	-0.2	27.1
		LPR	59.0	41.0	24.5	7.0	6.8	2.4	0.4	0.0	-0.2	26.0
		AKM	45.8	54.2	41.4	11.0	1.5		0.4	0.0	-0.2	3.6
		BS										37.3
	10-15 yrs	true	48.9	51.2	34.7	6.6	6.8	3.0	0.2	0.2	-0.3	22.4
		LPR	48.9	51.1	34.1	7.1	7.1	2.7	0.3	0.1	-0.2	22.8
		AKM	37.8	62.2	49.7	11.5	0.9		0.3	0.1	-0.3	1.9
		BS										37.1
	>15 yrs	true	37.5	62.6	43.6	6.1	7.8	5.3	0.1	-0.1	-0.2	24.0
		LPR	38.3	61.7	42.7	7.2	7.9	4.2	0.1	-0.1	-0.2	22.5
		AKM	30.0	70.0	59.1	12.2	-0.9		0.2	-0.2	-0.3	-1.7
		BS										39.5
1999-01	<5 yrs	true	58.3	41.7	20.3	6.6	8.0	3.4	1.4	1.3	0.7	34.4
		LPR	58.5	41.5	19.8	7.0	7.4	3.6	1.8	1.0	0.9	31.4
		AKM	46.4	53.6	35.7	10.1	4.4		1.3	1.2	0.8	11.6
		BS										40.8
	5-10 yrs	true	55.1	44.9	25.5	5.9	8.4	2.4	1.1	1.2	0.5	34.1
		LPR	55.4	44.6	25.3	6.0	8.0	2.3	1.4	1.0	0.6	32.5
		AKM	43.9	56.1	39.1	9.3	5.0		1.1	1.2	0.5	13.0
		BS										41.4
	10-15 yrs	true	50.7	49.3	29.6	6.5	8.7	2.4	1.0	0.8	0.3	31.4
		LPR	50.9	49.1	29.8	6.6	8.1	2.3	1.1	0.8	0.4	29.0
		AKM	40.4	59.6	43.2	9.9	4.3		1.0	0.9	0.4	10.4
		BS										42.8
	>15 yrs	true	49.0	51.0	33.2	5.6	6.8	3.3	1.2	0.7	0.3	24.8
		LPR	49.3	50.7	33.4	5.5	6.7	2.8	1.3	0.6	0.3	24.9
		AKM	39.4	60.6	47.5	10.7	0.3		1.1	0.8	0.3	0.6
		BS										36.4
2011-13	<5 yrs	true	36.9	63.2	41.4	5.6	7.4	8.0	3.0	-1.9	-0.4	24.2
		LPR	38.8	61.2	40.6	5.9	6.7	7.2	2.6	-1.5	-0.2	21.6
		AKM	33.4	66.6	58.3	10.0	-2.2		4.5	-4.7	0.8	-4.6
		BS										44.9
	5-10 yrs	true	43.9	56.1	34.0	5.6	7.5	8.1	1.6	-0.6	-0.1	27.4
		LPR	45.0	55.0	33.4	6.1	7.6	7.0	1.6	-0.6	-0.1	26.8
		AKM	33.0	67.0	54.9	11.3	0.7		1.4	-1.2	-0.2	1.4
		BS										39.2
	10-15 yrs	true	37.4	62.6	35.5	13.0	8.9	5.2	1.0	-0.8	-0.3	20.7
		LPR	38.2	61.8	35.0	13.3	8.5	5.1	0.9	-0.7	-0.2	19.6
		AKM	33.0	67.0	54.9	11.3	0.7		1.4	-1.2	-0.2	1.4
		BS										54.7
	>15 yrs	true	42.2	57.8	38.4	6.1	8.9	4.1	0.8	-0.5	0.0	28.9
		LPR	42.6	57.4	38.1	6.6	8.6	3.8	0.8	-0.5	0.0	26.9
		AKM	34.4	65.6	53.4	11.3	0.5		1.0	-0.6	-0.1	1.0
		BS										47.5

Notes: The simulations involve 1,000,000 workers and 100,000 firms where $K = 14, L = 24$. The length of the panel is 7 years. We run 10 simulations from the estimated model in steady state. “Model = true” is the infeasible variance decomposition as if heterogeneity were observed. “Model = AKM” refers to (therefore misspecified) AKM estimates. “Model = LPR” refers to our model. “Model = BS” refers to Borovičkovà and Shimer (2017)’s calculation of the degree of sorting. We show the mean simulation. With such a big sample size, there is very little variance across simulations.

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.[17]

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.

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

[17] 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).

(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,

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

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

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

where the subscript indicates which variable indexes rows and which variable indexes columns. It follows from (13) 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:

$$p(z,\ell,w,\ell') = \sum_k p(z|k)\,p(k,\ell)\,p(w|k,\ell)\,p(\ell'|k,\ell), \tag{14}$$

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

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

with $p(k,\ell,\ell') = p(k,\ell)\,p(\ell'|k,\ell)$. Equation (14) 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 $A_1 = P_{\ell 1},\ldots,P_{\ell L}$, 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 $U_\ell^\top P_{\ell\ell'} V_\ell \Lambda_\ell^{-1}, \ell' = 1,\ldots,L$, can be simultaneously diagonalized. The matrices of eigenvalues are $\Delta_{\ell\ell'}\Delta_\ell^{-1} = \mathrm{diag}[p(\ell'|k,\ell)]$, 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).[18]

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':

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

Denoting $\Delta_{\ell\ell'} = \mathrm{diag}[p(k,\ell,\ell')]$, 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 $\frac{p(\ell|k,\ell)}{p(\neg|k,\ell)} \neq \frac{p(\ell|k',\ell)}{p(\neg|k',\ell)}$.*

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

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 $\overline{\Delta}_\ell^{-1} \Delta_{\ell\ell} = \mathrm{diag}[p(\ell|k,\ell)/p(\neg|k,\ell)]$ 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 $\mathrm{diag}[p(k,\ell)] = \overline{\Delta}_\ell + \sum_{\ell'} \Delta_{\ell\ell'}$ 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 $\Delta_\zeta = \operatorname{diag}[q(\ell, \zeta)]$. Because $q(\ell) = \sum_\zeta q(\ell, \zeta) \neq 0$ by assumption, the identification lemma identifies $\Delta_\zeta \Delta_{\zeta_0}^{-1} = \operatorname{diag}[q(\zeta|\ell)]$. 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,

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

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 r_1j / r_1i ∉ {r_ij, 1/r_ij} then r_1j ← 1/r_1j
        If r_1j / r_1i = 1/r_ij then r_ij ← 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

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

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 $\widetilde{\nu}_\ell \propto \nu_\ell \gamma_\ell$. For each solution $(\lambda_\ell, \nu_\ell, \gamma_\ell)$ there is an observationally equivalent one $(\widetilde{\lambda}_\ell, \widetilde{\nu}_\ell, \widetilde{\gamma}_\ell) \propto \left(\frac{\lambda_\ell}{\gamma_\ell}, \nu_\ell \gamma_\ell, \frac{1}{\gamma_\ell}\right)$, 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

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

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.[19]

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

[19] 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

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

notations, for $\ell = 1, ..., L$,

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

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

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

subject to the normalization $\sum_{\ell=1}^{L} \gamma_{k\ell} = 1$.[20] 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 $\ell = 1, ..., L$,

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

[20] 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.

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

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

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

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

$$\delta_{k\ell}^{(s+1)} = \left(1-\lambda_k^{(s+1)}\right)n_{k\ell 0}\left[n_{k\ell\neg}\frac{1-\delta_{k\ell}^{(s)}-\lambda_k^{(s)}}{\overline{M}_{k\ell}^{(s)}} + n_{k\ell 0}\right]^{-1}.$$

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

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

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, $(\psi^{(m+1)}, \delta^{(m+1)}, \lambda^{(m+1)}, \nu^{(m+1)}, \gamma^{(m+1)})$.

E Numerical implementation

The implementation of the estimation allows the estimation to be scaled up to larger data sets by expansion of the number of CPUs in the computing cluster. The following describes how the storage and computation requirements of the estimation are delegated across CPUs in a parallel computing environment. The coding is done in Fortran and parallelization is performed with OpenMPI.

E.1 Data structure

The Danish Matched Employer-Employee (MEE) data comprise $I = 4,000,000$ workers and $J = 400,000$ firms observed at a weekly frequency from 1985 to 2013. The fundamental observation in the data is a spell (either employment or non-employment).

A worker history consists of a series of employment and unemployment spells. It is stored as a linked list. Each object in the list is a spell. The spell object contains,

- Start and end weeks of the spell.
- ID's of the worker and firm (unemployment has firm ID 0).
- A vector of wage observations for each year of the spell.
- Pointers to the previous and next spell in the worker's history.
- Pointers to the previous and next spell in the firm's spell list (unlike the worker's linked list, the firm list is not necessarily chronological).

In addition, the data structure holds the observable characteristics of each worker and firm separate from the list of spells. The worker i object holds the worker's observable characteristics (gender, education, birth year, year of entry into labor market, etc) as well as pointers to the first and last spells in the worker's labor history. The firm j object holds observable characteristics (public-private) and pointers to the first and last spell in its list of spells. The firm $j = 0$ list holds all the non-employment spells in the data.

The data storage is divided across CPUs so that each CPU holds its own subset of worker histories. Denote by ι_c the set of worker IDs assigned to CPU c. Each CPU holds the entire set of firms, but CPU c's list of employment spells in firm j consists only of those that are contributed by workers in the subset ι_c.

The Danish MEE data set is relatively small by international comparison (by the small size of the Danish population). Nevertheless, it does place significant demands on computer memory. Needless to say, this issue only becomes more acute for MEE data from larger countries. It is a virtue of the code that the memory requirement associated with each CPU is roughly $1/C$ of the total size of the data given a total of C CPUs. Thus, the memory pool available to the estimation is the combined memory of the nodes in the cluster, which is trivially scaled up by adding more nodes. This opposed to a data structure where each CPU holds the entire data set, which would place heavy memory requirements on multi-CPU nodes.

E.2 E-step

E.2.1 Likelihood evaluation for a given $(\beta, \mathscr{L})$.

Each CPU holds its own copy of the firm classification, $\mathscr{L}$. With this, CPU c evaluates $L_i(\beta, \mathscr{L}) = \sum_{k=1}^{K} L_i(k; \beta, \mathscr{L})$ for any $i \in \iota_c$ by walking through the worker i linked list of spells. CPU c calculates $L^c = \sum_{i \in \iota_c} \ln L_i(\beta, \mathscr{L})$. The likelihood of the data is then found by summing L^c across CPUs, $L(\beta, \mathscr{L}) = \exp\left(\sum_{c=1}^{C} L^c\right)$. This is a modest communication of a single double precision number across the C CPUs. The calculation of the overall likelihood is not necessary for the execution of the E-step, but serves as useful check that the algorithm is indeed proceeding to increase the likelihood in each iteration.

E.2.2 Worker posterior update for a given $(\beta, \mathscr{L})$.

CPU c updates worker posteriors for all $i \in \iota_c$ by, $p_i(k; \beta, \mathscr{L}) = L_k(k; \beta, \mathscr{L}) / L_i(\beta, \mathscr{L})$. No communication across CPUs is necessary for this and CPU c knows only the posteriors for workers $i \in \iota_c$. Nowhere in the CEM algorithm does CPU c need to know the worker posterior for workers outside ι_c. This is a significant savings in communication which would otherwise involve a communication of $I \times K$ double precision numbers across the C CPUs in each E step.

E.3 M step

The M step uses the updated posterior $p_i(k; \beta, \mathscr{L})$ from the E step. Each part of the M step requires only modest communication between nodes.

E.3.1 $\pi_k(z)$ update for given $(\beta, \mathscr{L})$.

With the worker posteriors in hand CPU c calculates $\pi_{k,c}(z) = \sum_{i \in \iota_c} p_i(k; \beta, \mathscr{L}) \mathbf{1}\{z_i = z\}$, which is communicated across the CPUs. This is a $K \times Z$ dimension double precision array communication across C CPUs where each CPU receives $\sum_{c=1}^{C} \pi_{k,c}(z)$.[21] Each CPU then calculates $\pi_k(z) = \sum_{c=1}^{C} \pi_{k,c}(z) / \left[\sum_{k=1}^{K} \sum_{c=1}^{C} \pi_{k,c}(z)\right]$.

E.3.2 $m_{k\ell}(x)$ update for given $(\beta, \mathscr{L})$.

CPU c calculates $m_{k\ell,c}(x) = \sum_{i \in \iota_c} p_i(k; \beta, \mathscr{L}) \mathbf{1}\{x_{i1} = x, \ell_{i1} = \ell\}$, which is communicated across the CPUs with each CPU receiving $\sum_{c=1}^{C} m_{k\ell,c}(x)$. This is a $K \times L \times X_{ini}$ double precision array where X_{ini} is the number of x categories in the initial distribution. Each CPU then calculates $m_{k\ell}(x) = \sum_{c=1}^{C} m_{k\ell,c}(x) / \left[\sum_{\ell=1}^{L} \sum_{n=1}^{N} m_{k\ell,c}(x)\right]$.

[21]Using mpi_allreduce.

E.3.3 Wage parameters for given $(\beta, \mathscr{L})$.

CPU c calculates $\mu_{k\ell,c}(x) = \sum_{i\in\iota_c} p_i(k;\beta,\mathscr{L}) \sum_{t=1}^T \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} w_{it}$ and $d_{k\ell,c}(x) = \sum_{i\in\iota_c} p_i(k;\beta,\mathscr{L}) \sum_{t=1}^T \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}$. These 2 $K \times L \times X$ arrays are communicated across CPUs to form $\sum_{c=1}^C \mu_{k\ell,c}(x)$ and $\sum_{c=1}^C d_{k\ell,c}(x)$, where X is the number of relevant x categories for the wage parameters as well as γ and λ mobility parameters. Each CPU proceeds to calculate $\mu_{k\ell}(x) = \sum_{c=1}^C \mu_{k\ell,c}(x) / \sum_{c=1}^C d_{k\ell,c}(x)$.

Moving to the variance, CPU c calculates $\sigma_{k\ell,c}(x) = \sum_{i=1}^I p_i(k;\beta,\mathscr{L}) \sum_{t=1}^T \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}[w_{it} - \mu_{k\ell}(x)]^2$. The $K \times L \times X$ array is communicated across CPUs to form $\sum_{c=1}^C \sigma_{k\ell,c}(x)$. Each CPU calculates $\sigma_{k\ell}(x) = \sqrt{\sum_{c=1}^C \sigma_{k\ell,c}(x) / \sum_{c=1}^C d_{k\ell,c}(x)}$.

E.3.4 Mobility parameters for given $(\beta, \mathscr{L})$.

Running through worker spell lists, each CPU calculates mobility counts,

$$\overline{n}_{k\ell,c}(x) = \sum_{i\in\iota_c} p_i(k;\beta,\mathscr{L}) \#\{t : D_{it} = 0, \ell_{it} = \ell, x_{it} = x\}$$

and

$$n_{k\ell\ell',c}(x) = \sum_{i\in\iota_c} p_i(k;\beta,\mathscr{L}) \#\left\{t : D_{it} = 1, \ell_{it} = \ell, \ell_{i(t+1)} = \ell', x_{it} = x\right\}.$$

These two integer arrays (of size $K \times (L+1) \times X$ and $K \times (L+1)^2 \times X$, respectively) are communicated across CPUs to form $\overline{n}_{k\ell}(x) = \sum_{c=1}^C \overline{n}_{k\ell,c}(x)$ and $n_{k\ell\ell'}(x) = \sum_{c=1}^C n_{k\ell\ell',c}(x)$. With these counts each CPU updates $\gamma_{k\ell}(x)$, $\lambda_\ell(x)$ and $\nu_\ell(x)$ according to section D.

E.4 C step

The C-step reassigns firm types in such a way as to increase the value of the expected log likelihood function, thereby increasing the likelihood of the data. The C step can be viewed as a simple extension of the M step where the firm classification is just another set of parameters to be chosen so as to improve on the expected log likelihood. While the M step requires very modest communication, the C-step does involve J separate communications of size L arrays within the cluster. This is a significant communication load and consequently, it is advantageous to do multiple EM iterations between C steps.

The firm IDs have been chosen so that firms are ordered by size ($j = 1$ is the largest firm where size is the number of wage observations throughout the panel). The algorithm reassigns firm type j by,

$$\ell_j^{(s+1)} = \arg\max_\ell \sum_{i=1}^I \sum_{k=1}^K p_i(k; \widehat{\beta}^{(s)}, \mathscr{L}^{(s)}) \ln L_i(k; \widehat{\beta}^{(s)}, \mathscr{L}_{-j}^{(s)}(\ell)), \tag{15}$$

where $\mathscr{L}^{(s)}_{-j}(\ell)$ is the firm classification that is obtained by taking the $\mathscr{L}^{(s)}$ classification where all firm types $j' = 1, \ldots, j-1$ have already been reassigned, and furthermore replace the j'th element with ℓ. Do the reassignment in order. This step increases the expected log likelihood.

Done naively, the step is expensive since it involves $L \times J$ expected likelihood evaluations of the data. But the expected log likelihood varies with firm j's type only through the spells that directly involve firm j and through firm j's type's impact on the $q(\ell, \mathscr{L}^{(s)}_{-j}(\ell))$ distribution. The latter does involve all spells but in a way that allows simplification. Define by $\Omega(\mathscr{L})$, the contribution to the expected log likelihood from the $q(\cdot|\mathscr{L})$ related terms,

$$\Omega(\beta, \mathscr{L}) = -\sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k; \beta, \mathscr{L}) \left[\ln q(\ell_{i1}|\mathscr{L}) + \sum_{t=1}^{T} D_{it} \ln q(\ell_{i(t+1)}|\mathscr{L}) \right]$$

Define,

$$n^q_\ell(\mathscr{L}) = \sum_{i=1}^{I} \left[\mathbf{1}\{\ell_{i1} = \ell\} + \#\left\{t : D_{it} = 1, \ell_{i(t+1)} = \ell\right\} \right]$$

with which we can write,

$$\Omega(\beta, \mathscr{L}) = -\sum_{\ell=1}^{L} \ln q(\ell|\mathscr{L}) n^q_\ell(\mathscr{L}).$$

It is worth noting that another way of calculating Ω is by adding up spells at the firm level. Denote by $\hat{n}_j$ the number of employment spells in firm j,

$$\hat{n}_j = \sum_{i=1}^{I} \left[\mathbf{1}\left[j(i,1) = j\right] + \sum_{t=1}^{T} \mathbf{1}\left[D_{it} = 1, j(i,t+1) = j\right] \right].$$

with this, Ω can be written as,

$$\Omega(\beta, \mathscr{L}) = -\sum_{\ell=1}^{L} \ln q(\ell|\mathscr{L}) \sum_{j=1}^{J} \hat{n}_j \mathbf{1}\left[\ell_j = \ell\right] = -\sum_{\ell=1}^{L} \ln q(\ell|\mathscr{L}) \hat{n}(\ell|\mathscr{L}),$$

where the number of spells in type ℓ firms is,

$$\hat{n}(\ell|\mathscr{L}) \equiv \sum_{j=1}^{J} \hat{n}_j \mathbf{1}\left[\ell_j = \ell\right]. \tag{16}$$

This firm-centric formulation of Ω is the preferable one for the firm reclassification algorithm.

Continuing the firm-centric formulation of the log-likelihood, denote by $\iota(j) = \{(i,t) \mid j(i,t) = j\}$, that is, all worker-time pairs with firm j. We can then write the the firm j classification update as,

$$\ell_j^{(s+1)} = \arg\max_{\ell} \Bigg[\sum_{(i,t)\in\iota(j)} \sum_{k=1}^{K} p_i(k;\widehat{\beta}^{(s)},\mathscr{L}^{(s)}) \times \Big[f_{k\ell}(w_{it}|x_{it}) + (1-D_{it})\ln \overline{M}_{k\ell_{it}}(x_{it}) + D_{i(t-1)}\ln M_{k\ell_{i(t-1)}\ell} + D_{it}\ln M_{k\ell\ell_{i(t+1)}} \Big] + \Omega(\beta,\mathscr{L}_{-j}^{(s)}(\ell)) \Bigg]. \quad (17)$$

The algorithm is then as follows:

1. The firm j spell counts, $\hat{n}_j$, are determined at the outset of the overall estimation where all processors count how many spells they each have for each given firm j. $\hat{n}_j$ is then found by a communication of a size J integer vector across all processors. Furthermore, the firm IDs $j = 1,\ldots,J$, are ordered by firm size - specifically the size of $\iota(j)$. These steps are not done in the C-step but rather just once at the outset of the full CEM algorithm.

2. The firm classification at the outset of the C-step is $\mathscr{L}^{(s)}$. Denote by $\mathscr{L}^{(s),0} = \mathscr{L}^{(s)}$, where $\mathscr{L}^{(s),j}$ is the firm classification in the jth substep of the C-step. Initialize the C-step by the determination of $\hat{n}(\ell|\mathscr{L}^{(s)})$ by equation (16).

3. Take firm $j = 1$. Find the optimal firm type for firm j according to equation (17) and firm classification $\mathscr{L}^{(s),j-1}$. The (i,t) pairs in $\iota(j)$ are by the data delegation spread out across different CPUs. Each CPU evaluates the summation in equation (17) for its own (i,t) pairs for each firm type $\ell = 1,\ldots,L$. The data structure has for each firm defined a linked list of its spells held by CPU c, which allows quick within CPU evaluation of each CPU's contribution to equation (17). The full sum for each ℓ is then obtained by a summation across all CPUs to the master process. This is a communication of an L size array from each node to the master node. The master process resolves the maximization problem in equation (17), and communicates the optimal firm type $\ell_j^{(s+1)}$ to all CPUs, a single integer.

4. Update the firm classification $\mathscr{L}^{(s),j} = \mathscr{L}_{-j}^{(s),(j-1)}(\ell_j^{(s+1)})$. Thus, as the algorithm steps through $j = 1,\ldots,J$, the firm classification is updated sequentially with a new firm type for firm j. Also, update $\hat{n}^j(\ell) = \hat{n}(\ell|\mathscr{L}^{(s),j})$ and the type frequencies $q(\ell|\mathscr{L}^{(s),j})$. This is done by the simple algorithm (stated just for $\hat{n}^j$)

   (a) If $\ell_j^{(s+1)} = \ell_j^{(s)}$ then $\hat{n}^j(\ell) = \hat{n}^{(j-1)}(\ell)$ for all ℓ.

   (b) Else, $\hat{n}^j(\ell_j^{(s+1)}) = \hat{n}^{(j-1)}(\ell_j^{(s+1)}) + \hat{n}_j$ and $\hat{n}^j(\ell_j^{(s)}) = \hat{n}^{j-1}(\ell_j^{(s)}) - \hat{n}_j$. For all other firm types, $\hat{n}^j(\ell) = \hat{n}^{(j-1)}(\ell)$.

5. loop back to step 3 for next j. Exit when $j = J$ is completed. Denote by $\mathscr{L}^{(s+1)} = \mathscr{L}^{(s),J}$.

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002, March). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

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

Abowd, J. M., F. Kramarz, and S. Roux (2006, June). Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data. *Economic Journal 116*(512), 245–285.

Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock (2009, June). The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. In *Producer Dynamics: New Evidence from Micro Data*, NBER Chapters, pp. 149–230. National Bureau of Economic Research, Inc.

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.

Altonji, J. G., A. A. Smith, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*(4), 1395–1454.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(3), 673–697.

Bagger, J. and R. Lentz (2014, April). An Empirical Model of Wage Dispersion with Sorting. NBER Working Papers 20031, National Bureau of Economic Research, Inc.

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.

Borovičkovà, K. and R. Shimer (2017, November). High Wage Workers Work for High Wage Firms. NBER Working Papers 24074, National Bureau of Economic Research, Inc.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women *. *The Quarterly Journal of Economics 131*(2), 633–686.

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.

de Melo, R. L. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy 126*(1), 313–346.

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.

Eeckhout, J. and P. Kircher (2011). Identifying Sorting–In Theory. *Review of Economic Studies 78*(3), 872–906.

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

Gruetter, M. and R. Lalive (2009). The importance of firms in wage determination. *Labour Economics 16*(2), 149 – 160.

Guvenen, F., G. Kaplan, J. Song, and J. Weidner (2017, April). Lifetime Incomes in the United States over Six Decades. NBER Working Papers 23371, National Bureau of Economic Research, Inc.

Hagedorn, M., T. H. Law, and I. Manovskii (2017, January). Identifying Equilibrium Models of Labor Market Sorting. *Econometrica 85*, 29–65.

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.

Holzer, H. J., J. I. Lane, and L. Vilhuber (2004). Escaping low earnings: The role of employer characteristics and changes. *Industrial and Labor Relations Review 57*(4), 560–578.

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.

Iranzo, S., F. Schivardi, and E. Tosetti (2008). Skill dispersion and firm productivity: An analysis with employer-employee matched data. *Journal of Labor Economics 26*(2), 247–285.

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., C. Meghir, and J.-M. Robin (2016). Matching, sorting and wages. *Review of Economic Dynamics 19*, 63 – 87. Special Issue in Honor of Dale Mortensen.

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

Martins, P. S. (2008). Dispersion in wage premiums and firm performance. *Economics Letters 101*(1), 63 – 65.

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.

Woodcock, S. D. (2015). Match effects. *Research in Economics 69*(1), 100–121.