The Anatomy of Sorting – Evidence from Danish Data

Rasmus Lentz[*] Suphanit Piyapromdee[†] Jean-Marc Robin[‡]

October 9, 2022

Abstract

In this paper, we formulate and estimate a flexible model of job mobility and wages with two-sided heterogeneity. The analysis extends the finite mixture approach of Bonhomme et al. (2019) and Abowd et al. (2019) to develop a new Classification Expectation-Maximization algorithm that ensures both worker and firm latent type identification uses wage and mobility variations in the data. Workers receive job offers in worker type segmented labor markets. Offers are accepted according to a logit form that compares the value of the current job with that of the new job. In combination with flexibly estimated layoff and job finding rates, the analysis quantifies the four different sources of sorting: preferences (job values), segmentation, layoffs, and job finding. Job preferences are identified through job-to-job moves in a revealed preference argument. They are in the model structurally independent from the identified job wages, possibly as a reflection of the presence of amenities. We find evidence of a strong pecuniary motive in job preferences. While, the correlation between preferences and current job wages is positive, the net present value of the future earnings stream given the current job correlates much more strongly with preferences for it. This is more so for short than long tenure workers. In the analysis, we distinguish between type sorting and wage sorting. Type sorting is quantified by means of the mutual information index. Wage sorting is captured through correlation between identified wage types. While layoffs are less important than the other channels, we find all channels to contribute substantially to sorting. In early career, job arrival processes are the key determinant of both types of sorting, whereas the role of job preferences becomes increasingly important as cohorts age. Over the life cycle, job preferences intensify, type sorting increases and pecuniary considerations wane.

Keywords: Heterogeneity; Wage distributions; Employment and job mobility; Mutual information; Finite mixtures; EM algorithm; Classification algorithm; Sorting; Decomposition of wage inequality

JEL codes: E24; E32; J63; J64

[*]University of Wisconsin-Madison and Dale T Mortensen Centre, Aarhus University. 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 support from the Danish Council for Independent Research Social Sciences (grant DFF – 6109-00037) and the European Research Council (grant reference ERC-2016-CoG-682349).

[‡]Sciences-Po, Paris and University College London. Email `jeanmarc.robin@sciencespo.fr`

1 Introduction

To what extent are workers and firms sorted in the labor market? What drives sorting and how does it change over a worker's life? The answers to these questions are central to understanding sources of wage inequality and designing labor market policies. To this end, this paper estimates a flexible semi-parametric model of wages and mobility with two-sided unobserved and observed heterogeneity on a long matched employer-employee panel dataset. The broad-ranging interactions between tenure, experience and heterogeneity in wages and employment mobility that our model allows for, reveal considerable complexity in matching.

We propose a structural job mobility model where workers stochastically move in the direction of higher job values. Workers meet firms in labor markets segmented by worker type where both meeting rates and the distribution of vacancies by firm type are segment specific. In combination with flexibly estimated worker transitions in and out of unemployment, the mobility model allows us to quantify the impact of four different mobility channels on sorting: job values (preferences), segmentation, layoffs, and job finding. We do so through counterfactuals on the estimated model. While the layoff channel contributes less than the other channels, we find all channels to be significant contributors to sorting.

The estimated mobility model identifies a worker type's valuation of a match with a particular firm type relative to that of another firm type. This is separately identified from the match earnings. We find a significant positive correlation between a worker's valuation of the match and the associated current earnings. But more strikingly, when we calculate the net present value of the expected stream of future earnings given the current match, we find a substantially stronger correlation than that against current earnings. The correlation with net present value is quite strong, in particular at short tenure. Thus, we find convincing evidence of a strong pecuniary component in the variation of worker match valuations across firms. For the sake of emphasis: a worker's pecuniary compensation varies across firm types and the worker's revealed preferences show that the worker cares, particularly for short tenure workers. Furthermore, preferences contribute substantially to the sorting patterns we see in the data.

Our paper contributes to the worker-firm sorting literature by offering a rich and flexible framework where sources of sorting can be estimated. The framework allows for wage and non-wage attributes through both worker and firm heterogeneity. One big-picture message is that worker and firm heterogeneity condition mean wages and job preferences differently. Some firms are highly attractive, but they are not necessarily the ones paying the highest wages. Workers, on the other hand, need not share the same global ranking of job types. This is a rather different interpretation from what has been recently pushed forward by several authors (Lindenlaub, 2017; Lindenlaub and Postel-Vinay, 2020; Lise and Postel-Vinay, 2020; Lindenlaub and Postel-Vinay, 2021). In their interpretation, workers and jobs have multiple characteristics that interact in the match production function to generate different dimensions of sorting for each pair of worker-firm characteristics. They emphasize how difficult it then becomes to define a simple notion of sorting and to link wages and mobility to properties of the match surplus. However, they only consider observed attributes (O'NET) and use standard worker panel data (SIPP). Additionally, it is worth noting that the structure of our transition probabilities is perfectly consistent with search-matching theory, yet more flexible. Standard search-matching theory makes both mean wages and match preferences monotone functions of the same input, match productivity. But our model allows mean wages and match preferences to be separate parameters, rendered structurally independent possibly by the existence of amenities.[1] This story echoes Sorkin's (2018) recent paper, who first observes on US match employer-employee data that close to 50% of job-to-job transitions are not associated with an increase in earnings, and whose revealed preference estimation technique reveals that non-pay characteristics explain 2/3 of the variance of firm-level earnings and 15% of the overall earnings variance. Our paper is in a similar vein as Sorkin's (2018). An important difference is that his technique identifies commonly held firm rankings across workers, whereas the discrete-mixture approach we adopt identifies mean wages and job preferences that are both firm- *and* worker-specific. With this less restrictive setup, we can analyze and understand preference driven sorting outcomes. We can furthermore distinguish between sorting on wages and sorting on non-wage characteristics.

Since the seminal contribution of Abowd, Kramarz and Margolis (1999) (hereafter AKM) the literature studying how individual wages vary across workers and firms has focussed on estimating, by Ordinary Least Squares, a linear model with additive worker and firm fixed effects, and sometimes a match-specific effect.[2] As with any panel data model with limited mobility, the OLS estimator of the AKM fixed-effect model is prone to overfitting. In response, finite-sample bias corrections have been proposed (Andrews et al., 2008; Kline et al., 2020; Azkarate-Askasua and Zerecero, 2019). While they have provided a much needed correction to the bias in the framework's measurement of wage sorting, the AKM framework does not allow us to study the dynamics of sorting through workers' transitions between employment and unemployment, nor does it incorporate sorting that may arise from non-wage factors. To advance our understanding on this issue, we therefore depart from the usual fixed-effect estimation by adopting the recent approach of Bonhomme, Lamadon and Manresa (2019) (hereafter BLM) and Abowd, McKinney and Schmutte (2019), who regularize estimation by assuming a finite number of worker and firm types.[3]

[1] In all fairness, Lindenlaub and Postel-Vinay (2021) only consider employment and job transitions probabilities. Their notion of "worker-job surplus" is therefore independent of wages.

[2] 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).

[3] BLM put no shape restrictions on the way wages and job-matching depend on worker and firm heterogeneity, whereas Abowd et al. (2019) maintain AKM's additive structure of worker and firm effects

Our own model is more closely related to BLM in that we impose no restrictions on how wages depend on worker and firm latent types, and we use the Expectation-Maximization (EM) algorithm for estimation. Yet, we extend BLM's methods along three dimensions. First, BLM classify firms only on wages using a k-means algorithm and separately from the estimation of the rest of the model, including latent worker types. Instead we make firm classification a component of the main estimation procedure through a Classification EM algorithm (CEM) inspired by Celeux and Govaert (1992). Monte Carlo simulations show that our CEM algorithm improves over a one-step k-means. From the perspective of advancing our understanding of sorting and wage inequality, using mobility data to classify employers is crucial as two firms offering an identical wage but non-identical employment stability would result in different lifetime earnings for workers. Further, if workers have preferences for job stability then employer types should be classified by mobility patterns in addition to wages. We show it is indeed the case that jobs with low layoff risk are preferred.

Second, while BLM focus on wage residuals, we include observables of both workers (gender, education, tenure, and experience) and firms (sector and industry) in our estimation. Since unobserved heterogeneity may not be orthogonal to observed heterogeneity, e.g. some worker latent types may have faster wage growth over tenure or experience than others, we allow for a flexible relationship between latent types and observed heterogeneity. In other words, we let latent types be the channel through which observed *and* unobserved heterogeneity jointly determine wage and mobility outcomes. Thus, the inclusion of observable heterogeneity contributes to the identification of latent types in our population. Specifically, we assume that wages are log-normally distributed with means and variances that depend unrestrictedly on (x, k, ℓ), where x is time-varying worker heterogeneity (tenure and experience categories), k is the latent worker type and ℓ is the latent firm type. Furthermore, we allow wage residuals to be autocorrelated.

Third, we depart from BLM's totally flexible mobility model by using a parametric model for job-to-job transitions that reduces the number of mobility parameters by a factor of 10. This is good for efficiency, but the main reason for constraining the transition probabilities is intelligibility. The transition probabilities are specified as the product of a job sampling probability and a choice probability. The choice probability is an increasing function of the ratio of two job preferences (one for the incumbent employer and one for the poacher) where the job preference parameter can flexibly vary by the (x, k, ℓ) combination. This allows (i) job ranking to differ across both worker types k and their time-varying characteristics x, (ii) job ranking to flexibly differ from wage ranking, and (iii) the interpretation of a standard on-the-job search model with random preferences for job types and worker-specific offer arrival rates. Empirically, the model allows us to quantify the importance of job sampling or *"chance"* relative to worker preferences for job

in wage means while adding a stochastic match effect correlated with worker and firm types, and with mobility.

types or *"choice"* at different career stages. The nonlinear specification does however pose a threat to estimation feasibility. We develop an MM algorithm (Hunter, 2004; Hunter and Lange, 2004) nested inside the M-step of the CEM algorithm to overcome this estimation difficulty.

We focus on Denmark in the period $1987 - 2013$ in our empirical application.[4] First we find a considerable degree of interaction of worker and firm heterogeneity in mean wages. In Denmark, worker and firm types do not determine wages additively —there is a considerable degree of non-linearity. Second, broadening our measure of sorting, we employ the concept of mutual information (MI) that measures the distance of the estimated match distribution to independent matching which can flexibly represent the dependence between worker and firm latent types independently of their wage effects. In support of Bagger et al. (2013), we find that sorting is moderately increasing over time in Denmark, where some is positive assortative on wages. However, the MI reveals that the increased importance of amenities in sorting patterns as workers age and select into long tenure relationships – this conclusion would have been the opposite had we used the correlation of wage fixed effects as the sorting measure. Consistently, our estimated job preferences revealed from job-to-job moves, also confirm that as workers age and tenure rises, preferences are increasingly shaped by non-wage attributes.

Our mobility model allows us to run counterfactual and cohort simulations that comprehensively uncover the key drivers of type sorting measured by MI as well as the classic wage sorting measured by the wage fixed effect correlation. We find that a cohort achieves a significant part of its sorting pattern, both type and wage, early on as it enters the labor market. Thenceforth during a worker's early career, sorting is primarily driven by complementarities between worker and firm types in the offer arrival processes while employed and unemployed. These two channels are apparent particularly in the wage fixed effects correlation and thus seem to drive a classical form of sorting via wage effects. It is also important to observe that the mobility patterns out of unemployment, like the initial match draws, are a significant contributor to sorting and not the kind of reset that is often found in standard job search models. However, later in life, workers' subjective rankings of jobs become the dominant determinant of matching. The role of job preferences becomes increasingly important as cohorts age, and it is what gives rise to the positive age trend in type sorting. Moreover, the role of job preferences is more apparent in the MI index than in the wage fixed effects correlation implying that non-wage attributes play a greater role in match preference determination later in life.[5]

Finally, we show that the various channels of sorting are not independent from each

[4]We divide the observation period into five time windows and estimate the model separately for each window. A worker's and a firm's types are constant within each window.

[5]The divergence between MI and wage fixed effect correlation could be a result of both non-wage factors and non-linearity in wage fixed effects. However, we show in Appendix D.1 that the contribution of non-linearity or match effects on wage variance do not vary much by experience. Hence, the divergence between MI and wage fixed effect correlation over workers' career is more likely due to non-wage factors.

other. Our model explicitly considers four drivers of sorting: job preferences, layoffs, market segmentation, and reemployment. While job preferences and market segmentation appear to have little complementarity, we find that other channels interact. In particular, there is a strong complementarity between job preferences and layoff channels for type sorting given that the most preferred jobs tend to be those that last longer. This re-emphasizes the importance of classifying employer and employee types using both wage and mobility data.

The layout of the paper is as follows: Section 2 describes the model and the parametric specifications, Section 3 explains the estimation procedure. Then, Sections 4 and 5 present the estimation results, Section 6 analyzes sorting, and Section 7 concludes.

2 The data and model

2.1 The data

We use the Danish matched employer-employee data from 1987–2013. Wages and employment mobility are reported at annual and weekly frequency, respectively. We divide the large panel into five smaller five-year panels opening different windows on the business cycle: 1989–93, 1994–98, 1999–2003, 2004–08, 2009–13. The first period was one of high unemployment in Denmark (9.5% in 1993). The next period contains a decline in unemployment (to about 5% in 1998). From 1999 to 2008, unemployment stayed low, and went up again after the financial crisis (7.8% in 2012). For each five-year panel, we use information from the two preceding years to distinguish between short and long tenure jobs in the stock of jobs at the beginning of each period. For example, we use information from 1987–1988 for the stock of jobs in 1989.

We restrict the sample to employment spells that start after individuals attain their, in retrospect, highest education level. We remove all spells that start after the individuals turn 50 years old and treat any spell with a positive wage as an employment spell. Time between two jobs is in our analysis referred to as unemployment and it should be understood to be a broad definition that includes events like extended sick leave, but not education or retirement. Although most other studies keep only full-time workers, we keep most workers in the analysis and explicitly model wages and mobility. We do this because workers can move between part-time and full-time employment and because treating part-time jobs as unemployment would overstate the share of transitions out of employment. This strategy is unlike those in studies that use the AKM approach, as the latter typically include only full-time workers, neglecting an explicit model of job mobility. Confounding part-time employment with unemployment appears to us more inaccurate than confounding it with full-time employment.

Workers are indexed by $i \in \{1, ..., I\}$ and firms by $j \in \{0, 1, ..., J\}$, where $j = 0$ marks unemployment. Each worker i is either drawn from the working-age population in the

Table 1: Share of employment by experience and tenure

Tenure	Experience	period 1	2	3	4	5
	< 5	0.15	0.14	0.16	0.18	0.23
short	5-10	0.11	0.12	0.13	0.14	0.12
(< 100 days)	11-15	0.08	0.09	0.11	0.11	0.08
	>15	0.19	0.20	0.23	0.24	0.17
	< 5	0.03	0.02	0.02	0.02	0.03
long	5-10	0.08	0.07	0.06	0.06	0.08
(> 100 days)	11-15	0.08	0.08	0.07	0.06	0.08
	>15	0.28	0.27	0.22	0.18	0.21

first year of panel window, or enters the panel in the first week of the first year following his or her last year of schooling. Individual trajectories $(w_{it}, j_{it}, x_{it})_{t=1}^{T}$ are recorded at a weekly frequency, where $j_{it} \equiv j(i,t) \in \{0, 1, ..., J\}$ is the employer's ID in week t, x_{it} are time-varying controls, and w_{it} is the log of the worker's weekly earnings at occurrence t. Note that although the number of repeated observations T varies across individuals, we adopt the simplified notation of a balanced panel.

There is only one payroll recorded for each employment spell within a year. For example, if a worker transitions from firm 1 to firm 2 after week 12 in a given year, there would be one observation for the first 12 weeks in firm 1 and another for the next 40 weeks in firm 2. If the worker remains in firm 2 for the following year, then we have only one observation for the annual pay in that year. We calculate average weekly earnings (total earnings divided by the number of weeks) instead of hourly wages because the data on hours may be excessively noisy.

The worker's time-varying characteristics x_{it} include the short/long tenure status and potential experience (time since graduation).[6] Short tenure in a job is defined to be less than 100 weeks of employment (or two years). The model also makes a distinction between short and long term unemployment and we maintain the language of tenure as a shorthand for duration in the unemployed state also. For unemployment, short tenure is defined to be less than 26 weeks (6 months). We divide experience into four groups: less than 5 years, 5–10 years, 11–15 years, and more than 15 years. Table 1 displays the distribution of workers' tenure and experience. As expected, most younger workers have short tenure. Notably, older workers are nearly equally split between holding short and long tenure jobs. This is important to note as one could think that more experienced workers would be keeping the same job until retirement, a feature one might expect in less mobile labor markets than the Danish.

For each worker i, we observe a set of time-invariant characteristics z_i^w that include

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

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.

For each firm j, we observe a set of time-invariant firm characteristics z_j^f that include the public/mixed/private status and the industry of the firm.[7]

2.2 The model

We assume that employers (firms) can be clustered into L different groups indexed by $\ell \in \{1, \dots, L\}$ and that workers can be clustered into K different groups indexed by $k \in \{1, ..., K\}$. The index ℓ_j is the type of firm j and k_i is the type of worker i. Unemployment is a specific employment type and is denoted by $\ell = 0$. Worker and firm type assignments are assumed to be fixed over the duration of the 5-year panel.

We commence by stating the sample likelihood for a given firm classification $F = (\ell_1, ..., \ell_J)$ assigning a type ℓ_j to each firm j in the sample. We denote

$$q(\ell|F) = \frac{\#\{j : \ell_j = \ell\}}{J}$$

as the share of type-ℓ firms given a firm classification. For notational simplicity, we write $\ell_{it} = \ell_{j(i,t)}$ and $z_{it}^f = z_{j(i,t)}^f$ for the type and the observed characteristics of the firm employing worker i in period t. We also let

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

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

For a value β of the parameters (which will be discussed shortly in Section 2.4) and a classification F of firms, the complete likelihood for one worker i conditional on the first observation of tenure and experience x_{i1} is

$$\begin{aligned} \mathcal{L}_i(\beta|k, F) = & \frac{m_0(k, \ell_{i1}|x_{i1})\, \pi^w(z_i^w|k)\, \pi^f(z_i^f|\ell)}{q(\ell_{i1}|F)} f_{static}(w_{i1}|k, \ell_{i1}, x_{i1}) \\ & \times \prod_{t=1}^{T-1} M(\neg|k, \ell_{it}, x_{it})^{1-D_{it}} \left(\frac{M(\ell_{i,t+1}|k, \ell_{it}, x_{it})}{q(\ell_{i,t+1}|F)} \right)^{D_{it}} \\ \times \prod_{t=1}^{T-1} & f_{dyn}(w_{it+1}|k, \ell_{i,t+1}, x_{i,t+1}, w_{it}, \ell_{it}, x_{it})^{1-D_{it}} f_{static}(w_{i,t+1}|k, \ell_{i,t+1}, x_{i,t+1})^{D_{it}}. \quad (1) \end{aligned}$$

[7] We produced our own aggregation of the available industry code, but after much work we have not been able to obtain stable and sensible results across panels. We will thus not make a comparison in this aspect across time.

The structure of this likelihood summarizes all of the model's assumptions and notations, as we now explain.

Initial condition. A worker enters the panel with experience and tenure, x_{i1}, which is conditioned upon. Initial experience and tenure determine a particular distribution of initial matches $m_0(k, \ell_{i1}|x_{i1})$. This initial dependence reflects the endogeneity of tenure and matching for the trajectories of workers who are initially drawn from the stock.

Among workers of type k, there is a particular distribution of gender and education $\pi^w(z_i^w|k)$. For simplicity, we assume that gender and education are independent of job tenure and the matching firm, given the worker's type. This conditional independence assumption is in principle innocuous if the number of unobserved types K is large enough. Similarly, we also assume that the distribution of observed firm characteristics $\pi^f(z_i^f|\ell_{i1})$ is only a function of firm types. This is exactly like incorporating into the likelihood the usual ex post tabulations of observed individual characteristics by predicted latent type.

Lastly, the initial employer's ID j_{i1} is drawn given its type with a probability proportional to the relative frequency of each type in the population of firms. We thus assume that each firm within a group is equally likely to be selected. Given that inflows to firms are proportional to firm size in our data, this implies in particular that firm sizes within a group should be similar, as we expect large firms to have a higher sampling probability than small firms. Our firm classification algorithm is based on this likelihood. The estimated firm classification will thus be performed subject to the uniform-sampling assumption. Again, there is no loss of generality because, in principle, the algorithm can choose to group firms of different sizes in different groups if size matters for sampling.

Job mobility. In every period, the worker changes employment status (employed or unemployed) and employer type with probability $M(\ell_{i,t+1}|k, \ell_{it}, x_{it})$. This probability is conditional on worker type k, employer type ℓ_{it}, tenure and experience x_{it}. We do not specifically condition on observed worker and firm characteristics. Since this is a clustering model, the group (k or ℓ) is a sufficient statistic for all socioeconomic determination. For instance, if gender per se were to determine mobility and wages, then the proportion of females would vary by group and show in probabilities $\pi^w(z_i^w|k)$.

The specific employer j_{it} is drawn inversely proportionally to the number of firms of type $\ell_{i,t+1}$, that is $1/q(\ell_{i,t+1}|F)$. This is a first difference to be emphasized with BLM. In BLM, the firm classification is operated ex ante and the sampling mechanism of firms given types can be neglected as it contains no parameter left to be estimated. On the contrary, our algorithm will eventually classify firms based on the same likelihood as the workers'. Firm types ℓ_j and the sampling probabilities $1/q(\ell_{i,t+1}|F)$ are therefore parameters as far as estimation is concerned.

The worker stays with the same employer with probability

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

Finally, we assume that we do not know whether a mobility occurs for the last observation period .

Wages. The initial wage w_{i1} is drawn from a static distribution $f_{static}(w_{i1}|k, \ell_{i1}, x_{i1})$. We also model the wage distribution after a job-to-job transition or out of non-employment using the same static distribution. However, we allow for autocorrelation in the case of stayers. This assumption is motivated by tests of autocorrelation of wages within and between spells in the Danish data.

Due to the annual frequency of observations on job payrolls, we incorporate just one wage likelihood contribution for each job spell in a year. For example, in

$$f_{dyn}(w_{i,t+1}|k, \ell_{i,t+1}, x_{i,t+1}, w_{it}, \ell_{i,t}, x_{it}),$$

$w_{i,t+1}$ refers to the unique wage observation for the first week of the current spell-year and w_{it} refers to the unique wage observation for the first week of the preceding spell-year. For ease of notation, we use the same time index t. Firm types are static within a given estimation window. Thus, within a job spell it follows that $\ell_{i,t} = \ell_{i,t+1}$. However, since our wage distribution formulation does not impose it, we write the general form.

2.3 Identification

BLM extend results in Hu and Schennach (2008) and Hu and Shum (2012) and show that at least four wage observations are sufficient to identify mixtures of dynamic models for matched employer-employee data. In Appendix A, we adapt and simplify BLM's proof in the context of our specific assumptions. Because each employer change resets the wage dynamic process, three wage observations, one before the change and two after, are sufficient for identification. We use five-year panels in estimation. So, in theory, we should be on the safe side.

The proof uses the same assumptions as BLM's. First, there must be enough observed differences across firms – statistics of wage distributions, firm size, entry and exit flows, etc. – to identify the firm classification irrespective of specific worker trajectories. Second, all transitions are realized with positive probability. For example, all workers can go from a type-1 firm to a type-2 firm and move back. It is possible to weaken this assumption, like BLM do with the "alternating cycles", but the identification argument is essentially the same. We also assume that the transition probabilities, and ratios of transition probabilities, and products of such ratios differ by type. Basically, this assumes that the bipartite graph of worker-firm matches suffices to identify the latent worker types independently of wage observations. Finally, wage densities should be linearly independent with respect to worker types. That is, the wage density for type 1 cannot be obtained as a linear combination of the densities for types 2 and 3. This is the non-parametric equivalent of the full column rank condition of the regressor matrix in ordinary least squares.

2.4 Parametric specification

Although the identification proof is nonparametric, in practice we estimate a parametric model. The components of the parameter vector β are detailed herein.

2.4.1 Wage distributions

Static wage distributions (initial, out of unemployment, and upon employer change) are assumed lognormal given the match type. Specifically,

$$f_{static}(w|k,\ell,x) = \frac{1}{\omega_{k\ell}(x)}\varphi\left(\frac{w-\mu_{k\ell}(x)}{\omega_{k\ell}(x)}\right), \tag{2}$$

with $\varphi(u) = (2\pi)^{-1/2}e^{-u^2/2}$. This specification allows for a match-specific mean $\mu_{k\ell}$ and variance $\omega_{k\ell}^2$.

Within job spells (i.e. while $D_{it} = 0$) we assume serial residual correlation:

$$f_{dyn}(w'|k,\ell',x',w,\ell,x) = \frac{1}{\sigma_{k\ell'}(x')}\varphi\left(\frac{w'-\mu_{k\ell'}(x')-\rho\left[w-\mu_{k\ell}(x)\right]}{\sigma_{k\ell'}(x')}\right), \tag{3}$$

where (w',ℓ',x') are one period forward relative to (w,ℓ,x). This dynamic specification has the advantage that parameters $\mu_{k\ell}$ are mean wages. So we can still perform a variance decomposition as in the AKM literature. Note that f_{dyn} and f_{static} share the same mean $\mu_{k\ell}$ but have different variances $\omega_{k\ell}^2$ and $\sigma_{k\ell}^2$.

2.4.2 Mobility and preferences

The probability that a type k worker at time t transitions from a firm of type $\ell = 1,...,L$ to a firm of type $\ell' = 1,...,L$, is specified as

$$M(\ell'|k,\ell,x) = \lambda_{k\ell'}(x)P_{k\ell\ell'}(x). \tag{4}$$

Parameter $\lambda_{k\ell'}(x)$ is the worker type k conditional probability of meeting with a different employer of type ℓ'. The parameter $P_{k\ell\ell'}(x)$ is the probability that the meeting results in a transition from ℓ to ℓ' .

We assume a Bradley-Terry specification for $P_{k\ell\ell'}(x)$ (Agresti, 2003; Hunter, 2004).

That is,

$$P_{k\ell\ell'}(x) = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x) + \gamma_{k\ell'}(x)}. \tag{5}$$

Parameter $\gamma_{k\ell}(x)$, with $\sum_{\ell=1}^{L} \gamma_{k\ell}(x) = 1$, measures the perceived value of the match (k, ℓ, x). If the worker draws a same-type job, with no loss of generality, since $\lambda_{k\ell'}(x)$ is unrestricted, we assume that the worker moves with probability $1/2$.[8]

We emphasize in section 5 the interpretation that $\gamma_{k\ell}(x)$ is a monotone transformation of the value of match (k, ℓ, x). Sorkin (2018) is an example of such a view of job-to-job mobility. Search models such as Shimer and Smith (2000), Eeckhout and Kircher (2011), Lise et al. (2016), and Bagger and Lentz (2019) all offer variations on this theme. The virtue of our specification is to link the relative strength of preferences/values over the two matches, $\gamma_{k\ell}(x)$ and $\gamma_{k\ell'}(x)$, to the observed propensity of a worker type k's realization of a move from a type ℓ firm to a type ℓ' firm. That is, we assume that job-to-job mobility is a revelation of preferences over the two jobs involved while allowing for differences in chances to move, $\lambda_{k\ell'}(x)$. By not restricting the match value to be a function of mean wages $\mu_{k\ell}(x)$, we allow for the possibility that worker (k, x) may value more than just the wage in ℓ.

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

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

By convention, $M(0|k, 0, x) = 0$ since there is no transition from unemployment to unemployment. It follows that the probability of staying unemployed and the probability of staying with the same employer are

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

3 Estimation method

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 type is a random effect. Worker mobility across different firm types makes it impossible to separate the complete log-likelihood (i.e. $\sum_i \ln \mathcal{L}_i(\beta|k, F)$) across firm types. Consequently, the estimation delivers a point estimate for each firm type instead of a posterior probability distribution over it

[8]Parameters $\lambda_{k\ell'}, \gamma_{k\ell}$ are identified given knowledge of the unrestricted transition probabilities $M(\ell'|k, \ell)$. To wit, observe that $M(\ell|k, \ell) = \lambda_{k\ell}/2$ trivially identifies $\lambda_{k\ell}$ for all k, ℓ (and x), assuming no empty case. Then, choice probabilities $P_{k\ell\ell'}$ follow, and ratios $\gamma_{k\ell'}/\gamma_{k\ell}$ are identified by odds ratios given the normalization $\sum_{\ell=1}^{L} \gamma_{k\ell} = 1$.

as is the case for worker types. We therefore follow the approach in BLM of estimating a random effect model of wages and mobility given the firm classification by the EM algorithm. However, while BLM pre-estimate the firm classification using the k-means algorithm, we nest the worker EM algorithm inside a firm classification algorithm that we believe uses all the available information.

In this section, we first explain how we estimate β and posterior probabilities of worker types for a given firm classification F. We then explain how we set and update F. Lastly, we address the issue of the calibration of the number of groups K and L.

3.1 The EM algorithm for a given firm classification

For given likelihood parameters, β, and a firm classification F, the posterior probability that worker i is of type k is

$$p_i(k|\beta, F) = \frac{\mathcal{L}_i(\beta|k, F)}{\sum_{k=1}^{K} \mathcal{L}_i(\beta|k, F)}, \tag{6}$$

which is worker i's type k complete likelihood relative to the marginal likelihood of the observed data. Note that the factors $1/q(\ell|F)$ in the definition of $\mathcal{L}_i(\beta|k, F)$ in equation (1) appear in the numerator and the denominator of the posterior probabilities in the same way and can be simplified out. This means that given F, the posterior probability is unaffected by the firm sampling assumption.

The EM algorithm iterates the following steps:

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

M-step Determine the update $\beta^{(m+1)}$ as the β that maximizes the expected log-likelihood $\sum_i \sum_k p_i(k|\beta^{(m)}, F) \ln \mathcal{L}_i(\beta|k, F)$.

The M-step updating formulas for wage distributions are the usual posterior probability-weighted mean, variance and autocorrelation for Gaussian mixtures. For the dynamic specification, we can still exploit the model's linearity to derive analytical formulas.

M-step updating formulas for transition probability are simple frequencies in the unrestricted case. For job-to-job transitions, transition probabilities are nonlinear parametric specifications, and this poses an additional challenge. In Appendix B, we develop an MM algorithm (Hunter, 2004; Hunter and Lange, 2004) to maximize the expected log-likelihood subject to the parametric restriction on job-to-job transition probabilities $M(\ell'|k, \ell, x)$ very rapidly.[9]

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

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

3.2 Firm reclassification given other parameters

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

$$F^{(s+1)} = \arg\max_{F} \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k|\widehat{\beta}^{(s)}, F^{(s)}) \ln \mathcal{L}_i(\widehat{\beta}^{(s)}|k, F^{(s)}). \tag{7}$$

In practice, we only search for a firm reclassification that increases the likelihood i.e. some firms may keep the same types from the previous iteration if their types are already maximizing the likelihood. We first order firms by size. Then starting from the largest firm, say $j = 1$, we find $\ell_1^{(s+1)}$ such that it maximizes the criterion in equation (7), keeping all other firm types equal to their values in $F^{(s)}$. Then, we move to the second largest firm, say $j = 2$, and find $\ell_2^{(s+1)}$ given $\ell_1^{(s+1)}$, $\ell_3^{(s)}, ..., \ell_J^{(s)}$, and so on until the smallest firm, $\ell_J^{(s+1)}$. Thereafter, we return to the EM iterations with the updated $F^{(s+1)}$ as well as the updated firm types' sampling probabilities in the likelihood function.

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). Gibbs' inequality guarantees that the likelihood increases at each parameter update provided that one proceeds sequentially (sequential EM algorithm). Notice that, given L, and as shown by Celeux and Govaert (1992), the discrete classification should settle after a finite number of iterations. It should also be unaffected by very small changes in the other parameters. Hence, the asymptotic standard errors for β calculated given F should remain valid for the estimated classification. Of course, with administrative data, standard errors will be small thanks to the huge number of degrees of freedom.

3.3 Starting values and choice of group numbers K, L

The estimation uses twenty different starting values of parameters, and we choose the set of results with the highest likelihood value.[10] For the number of groups, we first cluster firms using the k-means algorithm: taking firms' wage quintiles, average size, inflow rate, and outflow rate as inputs of characteristics. We select the value of L associated with the highest Calinski-Harabasz index, which is the ratio of the between-cluster and the within-cluster sum-of-squares. The idea is that we want to choose an L that optimally represents distinctive groups of similar firms. We find the optimal L to be 14 for periods

Then, we 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

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

[10] We also checked results associated with the five best likelihood values and they are similar.

1–2 and 22 for periods 3–5.

In theory, one could apply the elbow method to pin down the number of worker types K. However, in practice, it is difficult to find an elbow in worker type clustering. Using AIC/BIC with the likelihood expression is also not possible when we have double-sided heterogeneity. Therefore, we choose the maximum number of K that we could handle computationally, specifically $K = 24$, and leave a theoretical exercise of pinning down K as future research.

4 What are these latent types?

In this section, we describe the characteristics of latent types and the distribution of matches. The estimation essentially gives us classifications of workers and firms. We assign labels to these groups in a way that we can relate our results to the literature and to understand how sorting may manifest through wage effects that are common across types. We label k and ℓ based on the global ordering of mean wages using a two-way fixed effects projection of the estimated mean wages $\mu_{k\ell}(x)$ with respect to the empirical distribution of match characteristics (k, ℓ, x) in the sample, i.e.

$$p(k, \ell, x) \propto \sum_{i,t} p_i(k) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}, \tag{8}$$

where the estimated posterior probability $p_i(k)$ is calculated using equation (6) at the estimated parameters β and F.[11] This projection takes the standard form:

$$\mu_{k\ell}(x) = \overline{\mu}(x) + a_k + b_\ell + \widetilde{\mu}_{k\ell}(x), \tag{9}$$

where $\overline{\mu}(x)$ represents the marginal effect of tenure and experience interactions, a_k is the worker effect, and b_ℓ is the firm effect.[12] The last term $\widetilde{\mu}_{k\ell}(x)$ is a residual capturing all remaining interactions. This "match effect" guarantees that the decomposition in (9) is an interpretation, not a restriction.[13] We relabel k and ℓ after estimation so that a_k and b_ℓ are now increasing in k, ℓ. Hence, by construction $\mu_{k\ell}(x)$ tends to increase with respect to indices k and ℓ. This relabeling renders an interpretation of higher-k and higher-ℓ as higher-wage types.

Note that $\mu_{k\ell}(x)$ corresponds to the expected current wage *unconditional* on past wages. The autoregressive coefficient ρ is estimated to be around $0.4 - 0.6$ across the five periods. This is not negligible but not huge. It therefore may mitigate a possible mis-

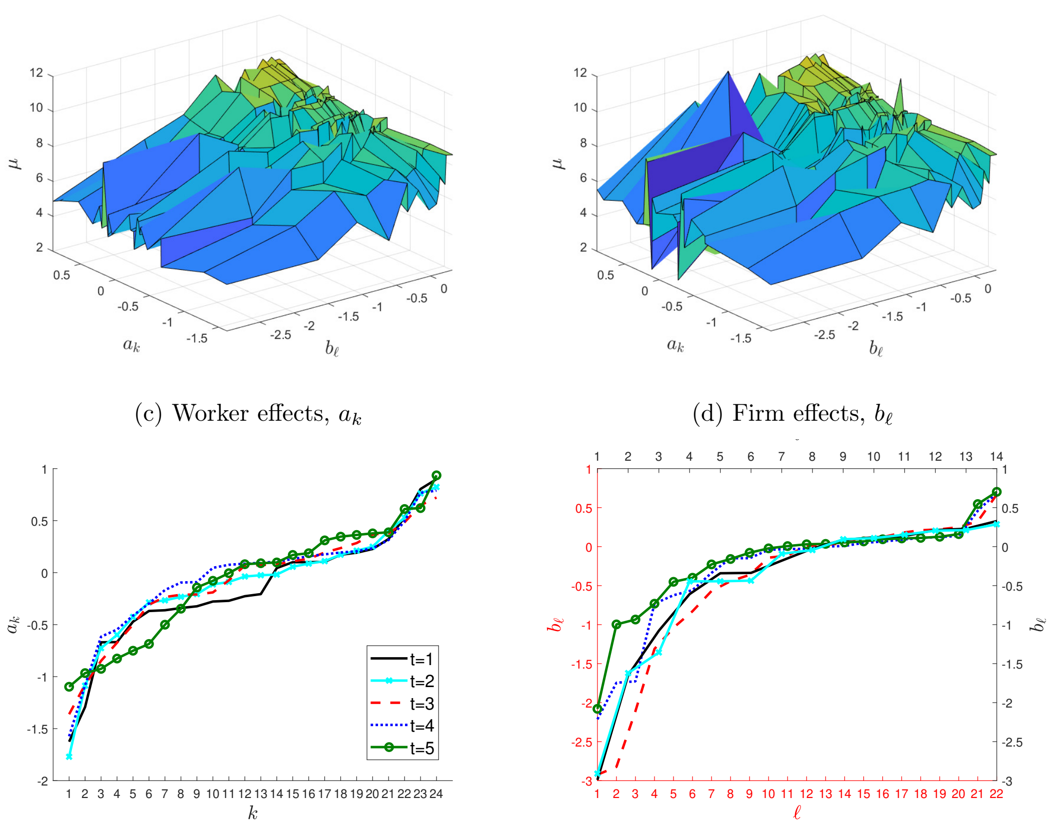
[11]The proportionality symbol $\propto$ means that the right-hand side of the "equality" needs to be normalized for the left-hand side to be a proper probability.

[12]The fixed effects are normalized to average to zero over observations.

[13]Note, that it is formally inappropriate to call it a "match effect". A proper match effect would require classifying matches (i, j) and not workers i and firms j separately.

Figure 1: Conditional mean wages (period 1: 1989-1993)

(a) Parameter $\mu_{k\ell}$ – Oldest, short tenure (b) Parameter $\mu_{k\ell}$ – Oldest, long tenure



Notes: Panels a and b show mean wages over worker and firm fixed effects by tenure. Short tenure in a job is defined to be less than 100 weeks of employment. Panels c and d display the worker and firm effects across five periods.

specification error in the assumed resetting property of employment and unemployment transitions.

4.1 Worker effects explain most of conditional mean wages

Figures 1a,b provide a graphical representation of the conditional mean wage $\mu_{k\ell}(x)$ for short and long tenure and the oldest worker group (the other age groups being similar) as estimated for the first period (1989–1993). We plot $\mu_{k\ell}(x)$ versus (a_k, b_ℓ) for given tenure and experience x. If $a_k + b_\ell$ was a good approximation, then the map $(a_k, b_\ell) \mapsto \mu_{k\ell}(x)$ would be a plane. At first sight, the linear approximation seems acceptable, although a careful reader might observe that along the a_k dimension, the graph is more horizontal at low values of b_ℓ than at higher values of b_ℓ.

Much of the literature on matched employer-employee data since AKM was concerned with log wage variance decomposition. Because this is not the may concern of this paper, we only here summarize the main results (see Appendix D for more details). The between-group variance – that explained by conditional means $\mu_{k\ell}(x)$ – explains about 50-60%, depending on the time period, of the total log-wage variance. Interactions between characteristics, absorbed in $\widetilde{\mu}_{k\ell}(x)$ account for about 25% of the between-group variance. Worker effects explain more than 40% of mean wages. Firm heterogeneity contributes a lot less to the log-wage variance than worker heterogeneity. Specifically, its contribution decreases over time from 14% in period 1 to 9% in the last period. Lastly, the covariance between worker and firm effects, a classic measure of sorting, explains a similar, modest portion of the between variance (about 10%).

Further, Figures 1c,d plot the "fixed effects" a_k and b_ℓ. Note that firm type ℓ has a different set of values in periods 1–2 ($L = 14$) than in periods 3–5 ($L = 22$). We display them as if they were different partitions of the same range. The two fixed effects seem to become flatter over time while keeping the same general shape. For workers, a_k shows greater dispersion, in particular at both extremes. For firms, such tail dispersion is only substantial for the lowest types. But most firm types have values of b_ℓ concentrated closer to zero. This explains why firm heterogeneity contributes little to the log-wage variance. However, while the firm type-specific wage premiums appear to have little dispersion, we will show in the later section that mobility heterogeneity leads to variation in net present values of future earnings associated with each (k, ℓ, x) match, and this plays an important role in determining workers' job preferences.

4.2 The worker index is monotone in education and tenure

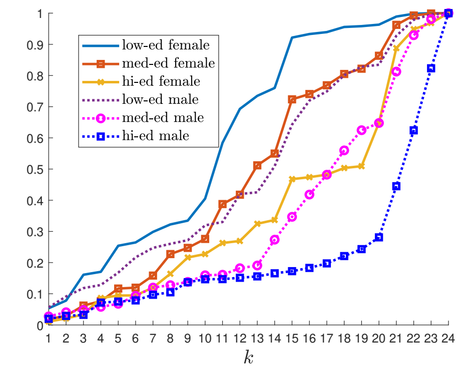
The marginal distribution of worker types k is close to uniform. Figure 2a shows the distribution of worker types conditional on gender and education $\pi^w(k|z^w)$ for period 1 (1989-93), with all periods producing very similar plots. Groups are clearly ordered by education and gender. There is a much greater proportion of low educated females in the smaller labels than in the greater ones. So worker types partially incorporate gender and education differences.

Figure 2b uses the estimated cross-sectional distribution $p(k, \ell, x)$ to partial out the distribution of worker types across experience. All CDFs across experience groups are close to each other, except for the CDF for the oldest workers (15+) which is slightly to the left of the other ones, indicating a tendency of older workers to be a bit less productive (assuming that wages reflect productivity).

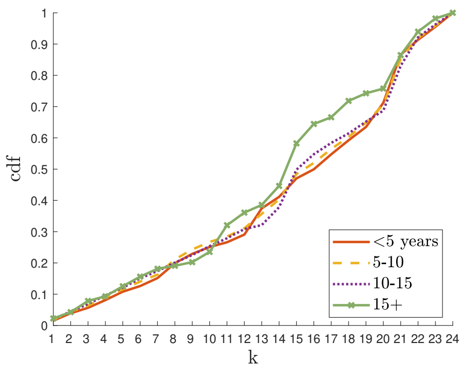
Figures 2c,d show the CDFs across tenure status while employed and unemployed. There is a clear stochastic ordering. Low-type workers tend to separate more from their employers than high-type workers. Low-type workers are also more likely to have long-term unemployment (greater than 26 weeks) than high-type workers. These patterns are robust across all five periods. So we conclude that tenure and worker types are correlated. Tenure is endogenous, and so we must allow transition probabilities to be tenure-dependent as how we have modeled it.

Figure 2: Characteristics of Worker Types (period 1: 1989-93)

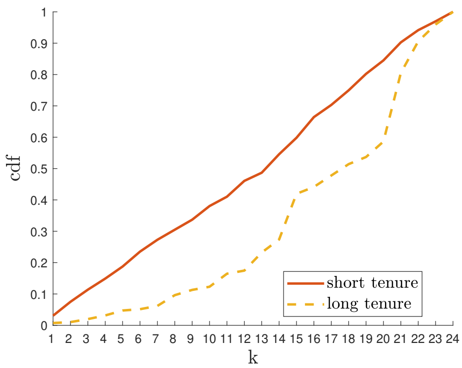
(a) By education and gender



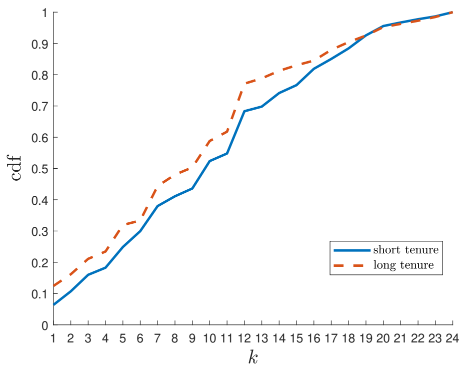
(b) By experience



(c) By employment tenure



(d) By unemployment tenure



Notes: CDF of worker type by education, gender and experience. Short tenure in a job is defined to be less than 100 weeks of employment. For unemployment, short tenure is defined to be less than 26 weeks.

Table 2: Firm Characteristics by Type (period 1: 1989-93)

ℓ	no. firms	no. workers	avg size	legal status private	public	mixed	avg inflow/yr	avg outflow/yr	avg age
1	8094	12127	1.50	0.62	0.34	0.04	0.64	0.62	3.69
2	5836	26324	4.51	0.66	0.32	0.02	0.54	0.51	6.01
3	15701	24080	1.53	0.86	0.10	0.04	0.71	0.70	3.04
4	2558	57614	22.52	0.74	0.23	0.03	0.54	0.44	6.51
5	26576	40283	1.52	0.90	0.08	0.02	0.46	0.45	4.74
6	12729	51080	4.01	0.84	0.16	0.00	0.43	0.41	7.49
7	12282	84309	6.86	0.93	0.02	0.04	0.53	0.49	4.02
8	91	371375	4081.04	0.15	0.57	0.27	0.33	0.02	8.80
9	4801	613119	127.71	0.81	0.16	0.03	0.37	0.22	7.79
10	21333	143988	6.75	0.87	0.12	0.00	0.34	0.31	8.05
11	18384	60604	3.30	0.97	0.01	0.02	0.56	0.53	5.66
12	9823	186200	18.96	0.95	0.01	0.04	0.44	0.36	5.15
13	32614	75256	2.31	0.91	0.07	0.02	0.36	0.35	5.24
14	27058	34746	1.28	0.92	0.04	0.04	0.68	0.67	3.32

4.3 Workers accumulate in firm clusters with lowest turnover

Table 2 contains descriptive statistics for firms in period 1, which is representative of all periods (see the online Appendix C for periods 2–5). Contrary to workers, the distribution of firm types is far from uniform. Firm groups starkly differ in firm sizes; the largest firms being clustered into smaller groups and the smallest in larger groups. Firm groups also vary in their turnover and age. Lower wage firms tend to have higher inflow and outflow rates, lower paid jobs tend to be less stable (e.g. Bagger and Lentz, 2019, and Jarosch, 2021). In terms of industry shares by firm type (not shown), public administration and public services are more prevalent in lower firm groups. The higher firm groups (offering higher wages) are dominant in the Construction, Transportation, and Communications industries in later periods. Finally, looking at all time periods (online Appendix), we see a (weak) declining trend of inflow and outflow rates by firm types over the whole estimation period. The share of firms with mixed public-private status has also increased.

For later reference, it is important to notice that two groups hire a majority of the workers (56%). Group 8 contains some very large firms from the public and mixed sectors. Group 9 also contains larger firms (more than 100 workers on average), mostly in the private sector. Both groups have relatively low exit rates, implying that their employees tend to stay employed longer. Firm clusters 10, 12 and 13 add another 23%. They are smaller private firms, also characterized by a small turnover. Similar patterns can be observed in all time periods.

Note that it is generally the case that entry and exit rates are close for all firm types. Firm clusters do not generally grow in size. Except, indeed, for the biggest groups, which have inflows significantly greater than outflows. The approach in this paper is not adapted to push this analysis further, because in particular we have chosen to estimate different models for different time periods. This modeling choice allows us to emphasize

Table 3: Classification stability of worker and firm latent types across periods

Period	Workers		Firms	
n to $n+1$	corr(k,k')	corr(a,a')	corr(ℓ,ℓ')	corr(b,b')
1 to 2	0.570	0.520	0.357	0.442
2 to 3	0.565	0.525	0.396	0.492
3 to 4	0.561	0.550	0.427	0.491
4 to 5	0.590	0.568	0.460	0.476

the features of labor markets that remain stable over the long term. However, it is clear that there is some significant non-stationarity in group sizes that will need to be properly addressed in future work.

4.4 Instability of latent groups across periods

In this section, we emphasize another type of non-stationarity. By estimating different models for different periods, worker and firm group labels may change arbitrarily across time. This may indicate a true change in the worker or firm types, or some statistical variance in the independent labeling. However, the amount of type change that we measure is suggestive of the existence of hidden dynamics in latent types.

In order to measure the stability of workers' labels across the different estimations, we assign to each worker in every period a type k_i corresponding to the highest posterior probability $p_i(k)$. Then we calculate the correlation between worker type in period n and worker type in period $n+1$ (see Table 3). We do so both in terms of the ordered groups k, as well as the associated a_k. These correlations are substantial (around 55 to 60%) while also demonstrating substantial individual type dynamics.

Table 3 also quantifies the stability of firm groups across periods. The correlation of firm types across periods is sizable (around 35 and 50%) but less than that of workers. We also reports firm group consistency in terms of the wage label b_ℓ, where correlations range between 45–50%. Note that the grouping stability analysis between periods 2 and 3 involves comparing a classification with 14 groups in period 2 with one that has 22 groups in period 3. Interestingly, it does not seem to affect the correlation between ordered group labels ℓ much.

4.5 Evidence of weak PAM

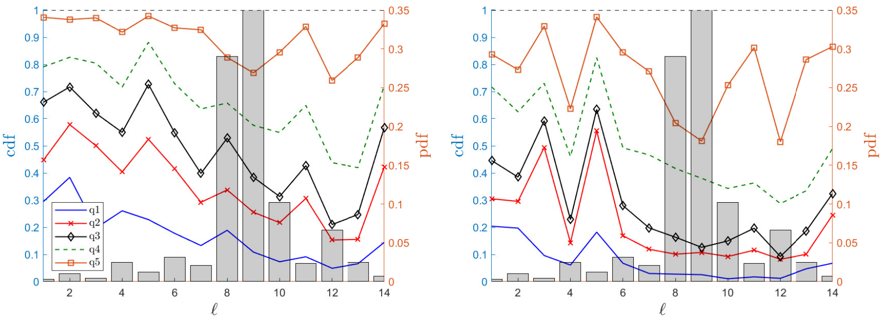
To illustrate the matching of workers to firms, we show how the worker type composition varies across firm types $p(k|\ell,x)$. For legibility, we plot the selected CDFs $\ell \mapsto p(1:k|\ell,x)$ as five vertically separated curves, for $k=4,8,12,16,20$, separately by tenure and experience (scale on the left vertical axis). We include the marginal distribution of firm types $p(\ell)$ (scale on the right vertical axis) with light gray bars.

Figure 3 shows the matching distributions for period 1, the other periods being similar.

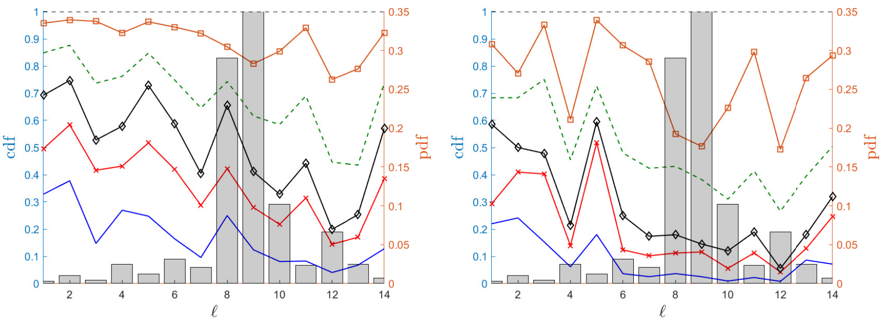
Figure 3: The distribution of workers by firms, $p(k|\ell, x)$, in period 1

Short tenure Long tenure

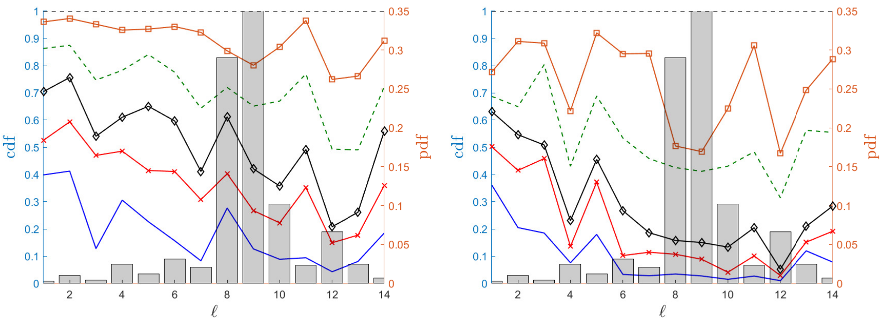
(a) Less than 5 years of experience



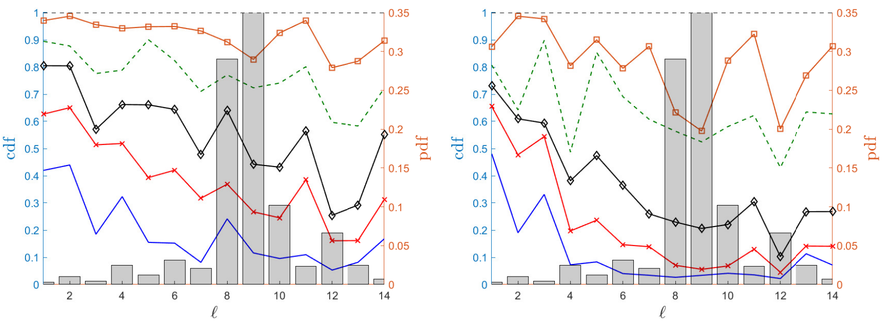
(b) 5-10 years of experience



(c) 10-15 years of experience



(d) at least 15 years of experience



Notes: The bar plots show the marginal distribution of firm types (scale on the right vertical axis), and the superimposed line plots show worker type composition across firm types where for legibility we combine worker types into five groups (sale on the left vertical axis).

It is clear by the relative increase in high type worker frequency in high type firms that there is positive sorting (Positive Assortative Matching, PAM) on worker and firm types. Sorting is nevertheless imperfect, and many workers – as it has already been emphasized – concentrate in a few firm types. This brings us to the next question of what determines job preferences.[14]

5 Job preferences

A worker's move from one job to another can be seen as an expression of revealed preference of the new match over the old. A number of papers have studied revealed preferences through job-to-job mobility such as Sorkin (2018), Bagger and Lentz (2019), Taber and Vejlin (2020) and Lentz et al. (2022).[15] Wages could indeed be a bad measure of match value in presence of compensating differentials. This is why, in our model, we have adopted the assumption that preferences (parameter γ) are independent from wage levels (parameter μ), and identification manifests itself in workers' mobility patterns. In addition, our model allows for the existence of heterogeneity in the probability of drawing a job offer (parameter λ). Finally, workers face idiosyncratic layoff (δ) and reemployment (ψ) probabilities. We shall study the role of unemployment shocks in the next section. We focus in this section on job-to-job mobility.

Our model is not an equilibrium model. A few lessons can nevertheless be drawn from the theoretical matching literature. In the frictional descendants of the partnership model of sorting in Becker (1973) such as Shimer and Smith, 2000, Gautier et al., 2010, Eeckhout and Kircher (2011) and Lise et al. (2016), an equilibrium with sorting is characterized by preference rank variation, and often disagreement across workers. In these models, workers climb and fall down their respective firm ladders at the same pace independent of the value differences across different rungs on the ladder. The impact of job-to-job mobility on match allocation depends purely to the ordinal properties of preferences.[16] Sorting in these models is a result of different workers ranking firms differently in equilibrium. Two worker types that have the same ordinal preferences will have the same equilibrium match allocation regardless of any cardinal differences.[17] These models are distinct in that

[14] The discrepancy is there because the number of spells per firm type shown in Table 2 includes all workers in our age range and was computed by counting all spells at each firm. On the other hand the pdf in Fig 3 was computed using posterior probability across k in each firm – workers who only have observations in the initial two-year periods will have zero posterior probability and hence not counted $p(k, \ell)$ in Fig 3.

[15] For earlier references, see also Villanueva (2007) and Usui (2008).

[16] Hornstein et al. (2011) emphasize this point in their discussion of job-to-job mobility rates.

[17] For example, in a positively assorted equilibrium, a low type worker may reject a match with the most productive firm even though, possibly, this could be the most productive the worker would be. However, in order to make the high type firm willing to forego the opportunity to match with a higher type worker, the low type worker would have to compensate the high type firm with an unacceptably low wage. The assumption in the model that the firm forfeits the opportunity to match with another worker for the duration of the current match is in this context essential for sorting. Postel-Vinay and Robin (2002) is an example where the elimination of this assumption in an otherwise similar mobility

fundamental complementarities in match values manifest in ordinal preference variation across agent types.

In contrast, the sorting in Lentz (2010) and Bagger and Lentz (2019) is a sorting equilibrium where worker types agree on the ranking of firms, but preference/match value intensity varies across workers in relation to expected search gains. So, preferences vary across types only in a cardinal sense. Sorting results from the worker's search intensity choice: the greater the value difference between origin and expected destination firm, the more intensely the worker searches away from the origin firm, and as a result generates a greater propensity to move.

Our model is a reduced form model insofar as it is agnostic as to how job preferences γ relate to match production and wages μ. It is also agnostic as to how job offer probabilities λ (or market segmentation) relate to preferences.

We end this discussion by showing how sorting in our model relates to the standard supermodularity property of match values. Think of $\gamma_{k\ell}(x)$ as a monotone transformation of the value of a (k, ℓ, x) match. Different worker types may face different ladders, i.e. different orderings of job types. Consider a worker in match (k, ℓ, x) receiving a job offer of type ℓ'. The probability of accepting the new job is $P_{k\ell\ell'}(x) = \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell'}(x)+\gamma_{k\ell}(x)}$ if mobility incurs a logistic mobility cost or amenity. Now, pick two worker types $k' > k$ and two firm types $\ell' > \ell$ such that $\frac{\gamma_{k'\ell'}(x)}{\gamma_{k'\ell}(x)} > \frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x)}$. Then, $P_{k'\ell\ell'}(x) > P_{k\ell\ell'}(x)$. Even if the two worker types agree on the ranking of ℓ versus ℓ', worker k' will climb her job ladder faster than worker k. Such ordering of the odds ratios $\frac{\gamma_{k\ell'}(x)}{\gamma_{k\ell}(x)}$ holds when the function $\ln \gamma_{k\ell}(x)$ is supermodular in (k, ℓ).

Of course, if worker k receives job offers at a lower frequency than worker k', this conclusion can be reversed. Understanding sorting therefore involves a complete analysis of the mobility parameters and their interactions. A particularly important feature of our model is that the "chance" component of transition probabilities, $\lambda_{k\ell'}(x)$ – that we call market segmentation – is separately identified from the "choice" component, $\gamma_{k\ell'}(x)$.

Absent of considerations regarding job offers, and assuming that workers start at the bottom rung, log supermodularity in γ will generate positive assortative matching. Note that this form of sorting implies a natural ordering of firm types that every worker recognizes and accepts. In reality, things could be a lot more complex. We could have different groups of workers with different ladders, for example. This is why, in the next subsection, we focus on detecting what we call "preference intensity", that is how far from indifference to firm types $\ell = 1, ..., L$ a worker (k, x) can be.

5.1 Job preferences intensify with increased ability and tenure

Indifference is maximal when $\gamma_{k\ell}(x) = 1/L$ for all ℓ. Note also that $\gamma_{k\ell}(x) \geq 0$ and normalization $\sum_\ell \gamma_{k\ell}(x) = 1$ allow to treat $\boldsymbol{\gamma}_k(x) := (\gamma_{k\ell}(x))_\ell$ as a discrete probability

environment results in no sorting even if there are complementarities in joint match values.

Table 4: Preference information by worker type, $d_{KL}\left(\boldsymbol{\gamma}_k(x)\,||\,\frac{1}{L}\right)$

		Short tenure				Long tenure			
Experience:		0-5	5-10	10-15	15+	0-5	5-10	10-15	15+
	Low	0.82	0.97	1.23	1.31	1.02	1.32	2.12	2.09
1989-93	Med	1.57	1.51	1.73	1.81	2.49	2.57	2.05	2.27
	High	2.26	2.09	2.15	2.64	2.39	2.25	2.34	2.69
	Low	0.86	0.78	1.07	1.17	2.42	1.72	1.86	1.64
1994-98	Med	1.40	1.36	1.50	1.86	1.96	1.45	1.63	2.02
	High	1.69	1.75	1.79	2.10	2.03	1.58	1.82	1.92
	Low	1.11	1.28	1.41	1.43	2.99	1.88	2.07	1.76
1999-03	Med	1.65	1.49	1.65	1.85	2.43	1.62	1.79	1.93
	High	1.94	2.06	2.31	2.62	2.37	2.32	2.23	2.68
	Low	1.13	1.28	1.28	1.40	3.21	2.59	2.39	1.86
2004-08	Med	1.88	1.79	2.13	2.12	2.65	2.31	2.77	2.36
	High	1.99	1.94	2.25	2.29	2.47	2.11	2.46	2.36
	Low	1.62	1.22	1.40	1.55	2.30	2.66	2.58	2.39
2009-12	Med	1.82	1.77	1.82	1.93	2.34	2.27	2.41	2.63
	High	2.44	2.40	2.55	2.62	2.84	2.80	3.00	3.09

distribution over ℓ. Therefore, for each (k, x), we calculate the Kullback-Leibler (KL) divergence from uniformity,

$$d_{KL}\left(\boldsymbol{\gamma}_k(x)\,||\,\frac{1}{L}\right) = \sum_{\ell=1}^{L} \gamma_{k\ell}(x) \ln\left(\frac{\gamma_{k\ell}(x)}{1/L}\right) = \ln L + \sum_{\ell=1}^{L} \gamma_{k\ell}(x) \ln \gamma_{k\ell}(x),$$

which, by Gibb's inequality, is non-negative, and is equal to zero when $\gamma_{k\ell}(x) = 1/L$ is uniform. Another interpretation is that $\ln L - d_{KL}\left(\boldsymbol{\gamma}_k(x)\,||\,\frac{1}{L}\right)$ is the Shannon entropy of distribution $\boldsymbol{\gamma}_k(x)$ given (k, x). The maximum entropy, or maximum surprise, is attained for the uniform distribution. On the other hand, the greater $d_{KL}\left(\boldsymbol{\gamma}_k(x)\,||\,\frac{1}{L}\right)$, the more intense is the preference for certain firm types instead of others. We here use "intensity" as an antonym for "indifference".

Table 4 shows $d_{KL}\left(\boldsymbol{\gamma}_k(x)\,||\,\frac{1}{L}\right)$, averaging over worker types k within three groups (low, medium, and high). If $d_{KL}\left(\boldsymbol{\gamma}_k(x)\,||\,\frac{1}{L}\right) = 0$, then we have complete indifference and in the language of labor search models we say that in this case there is no firm ladder for worker type k to climb. Evidently, workers face a ladder in the sense that they are not indifferent. We furthermore see a robust pattern that preferences for particular firm types strengthen in tenure. We also see such strengthening by experience for short tenure. The link with experience at long tenure is less clear.

There is also a strongly increasing relationship between job preferences and worker types, indicating that high-ability workers see greater value loss from mismatch. As explained above, the greater the expected gains, the more likely the move. This creates a

Table 5: Correlations between parameters given (k, x)

	Short tenure				Long tenure			
Experience:	0-5	5-10	10-15	15+	0-5	5-10	10-15	15+
$\gamma_{k\ell}(x), \mu_{k\ell}(x)$	0.41	0.43	0.42	0.40	0.22	0.25	0.20	0.20
$\delta_{k\ell}(x), \mu_{k\ell}(x)$	-0.25	-0.31	-0.32	-0.27	-0.12	-0.22	-0.23	-0.14
$\lambda_{k\ell}(x), \mu_{k\ell}(x)$	0.13	0.13	0.12	0.10	0.15	0.20	0.22	0.21
$\psi_{k\ell}(x), \mu_{k\ell}(x)$	0.23	0.24	0.24	0.22	0.19	0.23	0.23	0.22
$\delta_{k\ell}(x), \gamma_{k\ell}(x)$	-0.45	-0.46	-0.42	-0.42	-0.14	-0.23	-0.24	-0.26
$\lambda_{k\ell}(x), \gamma_{k\ell}(x)$	0.10	0.06	0.03	-0.01	0.04	0.10	0.11	0.07
$\psi_{k\ell}(x), \gamma_{k\ell}(x)$	0.31	0.31	0.27	0.24	0.25	0.29	0.31	0.29
$\delta_{k\ell}(x), \lambda_{k\ell}(x)$	-0.14	-0.15	-0.12	-0.11	-0.11	-0.14	-0.14	-0.08
$\psi_{k\ell}(x), \lambda_{k\ell}(x)$	0.85	0.83	0.81	0.76	0.75	0.80	0.75	0.70
$\psi_{k\ell}(x), \delta_{k\ell}(x)$	-0.19	-0.21	-0.17	-0.12	-0.11	-0.15	-0.15	-0.09
$\ln \gamma_{k\ell}(x), NPV_{k\ell}(x)$	0.79	0.78	0.74	0.73	0.53	0.61	0.59	0.62

Notes: The correlations are calculated over firm types ℓ, then averaged over worker types k and time periods, for given values of tenure and experience x. We use uniform weights in the calculation of the correlations across firm types ℓ. Note that parameters $\theta_{k\ell}(x)$, where θ stands for any of the parameters in the table, and normalized parameters $\frac{\theta_{k\ell}(x)}{\sum_{\ell>0}\theta_{k\ell}(x)}$ produce identical correlations across firm types.

basis for sorting whereby higher type workers are more likely to move to their preferred firm types, all else equal. If those preferred firm types also tend to be characterized by higher wages, then this results in positive wage sorting driven by the intensity variation of job preferences across worker types, regardless of whether worker types agree on the ranking of firms.

5.2 Strong pecuniary motive at short tenure, less so at long tenure

We proceed to ask how workers' preferences over firm types are related to the estimated characteristics of a (k, ℓ, x) match. Table 5 shows correlations between model parameters across firm types, averaged across worker types and time periods – but these correlations do not vary much over time – by experience and tenure. As one would expect, matches with higher wages (μ) and lower layoff risks (δ) are more preferred. We also see that short-tenure preferences are more closely aligned (ie have higher absolute correlations) with wages and layoff than long-tenure ones. Yet, the KL distance to indifference of job preferences was shown to increase with tenure. This implies that job preferences become more intense as tenure increases but pecuniary considerations become less important.[18]

Job preferences γ and reemployment rates ψ are weakly correlated (30%), at both short and long tenure. Hence, layoff shocks are not sending the workers back to the first

[18]It is possible that job destruction is endogenous to the value of a match in which case the negative correlation between preferences and layoff risk reflects the reverse causality; that higher valued jobs are less likely to terminate. We consider both interpretations reasonable.

rung of their ladders, but they are nevertheless strongly mixing. Job preferences and the rate at which offers arrive λ are orthogonal, indicating little ability of workers to generate more preferred offers. Market segmentation thus tends to slow down the speed at which workers move up their value ladder. Interestingly, job offers during employment λ and reemployment rates ψ are nearly perfectly correlated (80%). This indicates that unemployment and market segmentation work together in ways that are not aligned with workers' preferences for the jobs. Then choice incorporated in γ mixes things up, as γ is orthogonal to λ. One possible interpretation is that unemployed workers do not exert their choice as much as currently employed ones. Finally, layoff rates δ are orthogonal to λ and ψ.

To conclude, the more preferred jobs are broadly the more remunerating and long lasting ones. Unemployment and market segmentation are mixing. There is supporting evidence for partially directed search, but a large measure of search remains undirected.

5.3 Job preferences well aligned with net present values, particularly at short tenure

The preceding subsection showed that job preferences γ correlate with wage levels μ, job destruction δ and reemployment rates ψ, but that γ and job-to-job offer rates λ are orthogonal. We therefore end this section by considering a more comprehensive measure of pecuniary value of a job than current wages – the net present value of future earnings given current (k, ℓ, x). The NPV is a simple way of aggregating wages and employment shocks in a single index.

For this, we simulate 20 year histories forward starting in a (k, ℓ, x) match using a 5% annual discount rate. We prioritize simplicity in the NPV calculation: $NPV_{k\ell}(x) = \mathbb{E}\left[\sum_{t=0}^{T} \beta^t \mu_{k\ell_t}(x_t)\right]$, where the period t firm type match and tenure-experience realization (ℓ_t, x_t) follows the estimated laws of motion in the model and initial condition, $\ell_0 = \ell$ and $x_0 = x$. Furthermore, we set $\mu_{k0}(x) = 0$ for weeks of unemployment. This corresponds to a hand-to-mouth worker with log utility and a close to zero unemployment benefit replacement rate. It has the virtue of excluding wage variance $\sigma_{k\ell}(x)$ as a confounding factor in the calculation. NPV reflects purely the estimated log wage averages and the mobility model parameters. We artificially maintain calendar time fixed in the sense that the NPV calculation for a particular time window panel uses its own mobility model for the entire 20 year horizon.

Table 5, last row, shows the average worker type correlation between job values and net present value of future earnings. We see a strong positive relationship between a worker's job match preferences and the net present value of future earnings associated with the match – and this relationship is stronger than that between preferences and wages or layoff risks. The NPV relationship is also stronger for short than long tenure, and the difference between short and long tenure groups is decreasing in experience. We take this as evidence that pecuniary considerations (labor earnings and their loss) are stronger determinants of job preferences for younger workers and become less important as workers age, although they remain substantial. Additionally, we find an increasing correlation between preferences and NPVs over time (not shown).

It is worth noting that our estimation imposes no particular relationship between job preferences, wages and layoff risk. The correlations in Table 5 demonstrate that workers make job-to-job moves that reveal a preference for matches with firm types where the net present value of future earnings is greater, both because of higher wage and lower layoff risk. We also find that preferences are more substantially shaped by pecuniary considerations for short tenure workers. As workers become long tenured, preferences become more intense and non-wage attributes increase in importance. While our results are supportive that workers' preferences for job types are closely aligned with pecuniary considerations, the correlations are sufficiently far from perfect that a substantial remaining residual can be attributed to amenities as emphasized in both Sorkin (2018) and Hall and Mueller (2018) (in addition to the idiosyncratic mobility cost/amenity responsible for the stochastic nature of the mobility decision).

6 Sorting

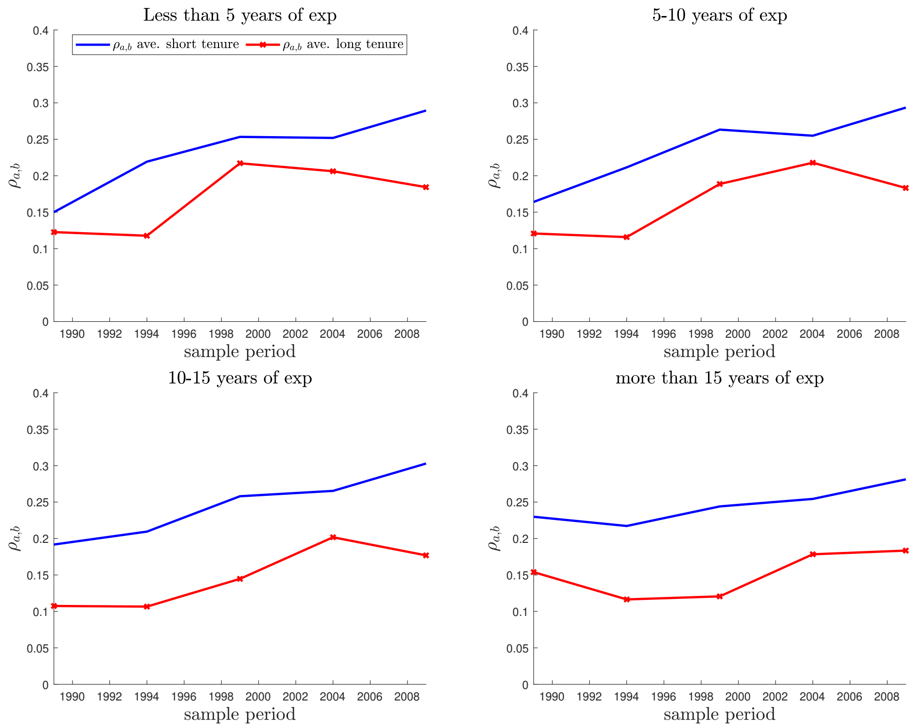
In this section, we measure sorting between latent worker and firm types. First, we do so by means of the correlation between the fixed effects obtained from the wage projection as typically done in the literature. Given our estimated importance of non-wage attributes in preferences as workers age and tenure rises, we then use a new measure of sorting, the mutual information, which allows us to flexibly account for both wage and non-wage sorting. Through the mutual information, we find a significant measure of sorting that is not reflected in the wage correlation measure. Furthermore, the estimated mobility model allows us to understand how sorting arises. In the last two subsections, we quantify the importance of different sorting channels and their interactions through counterfactual analyses.

6.1 Two measures of sorting

Wage sorting. The standard measure of sorting in the AKM literature is the correlation between worker and firm wage fixed effects. We show the correlation of a_k and b_ℓ conditional on tenure and experience in Figure 4 for matching from the cross-sectional distribution $p(k, \ell, x)$ in each period.

First, correlations increase with calendar time, which confirms results in Bagger et al. (2013) and Card et al. (2013) (on German data) although the increases we see are relatively modest. Second, comparing sorting by short and long tenure, the selection in jobs that last into long tenure reduces in wage sorting. This is in line with our results in section

Figure 4: Correlation by tenure and experience



Notes: correlations between worker and firm wage fixed effects by tenure and experience across five periods.

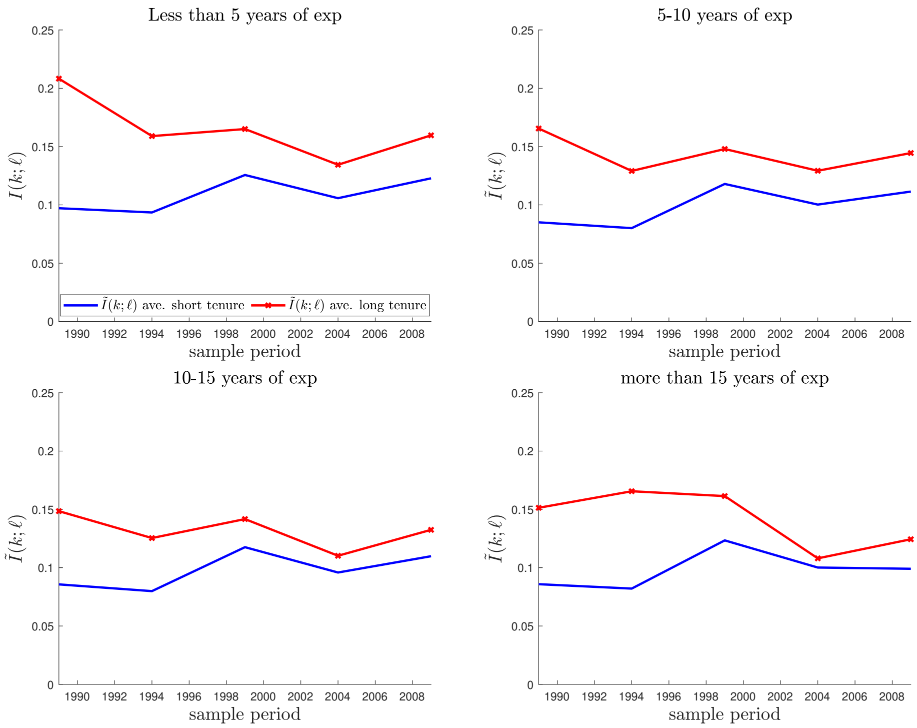
5 where we see substantially stronger dependence between wages and job preferences at short than at long tenure.

However, it is important to emphasize that the correlation of a_k and b_ℓ measures sorting on wage effects that are common across types. Sorting on wages that is not common across worker types will not be detected in the classic wage fixed effects correlation. Thus, the mapping of match preferences into wage fixed effects sorting is confounded by issues of search intensity, their commonality across worker types, and the importance of non-wage attributes.

Type sorting. There may be dependencies between worker and firm types that the correlation of fixed effects are missing. We already saw that worker and firm effects are not exactly additive in mean wages. Even if they were, there may be sorting patterns that are not aligned with wage attributes. The mutual information (MI), without imposing any structure, measures the dependence between two variables, say X and Y. It quantifies the information about one random variable through the knowledge about the other. Specifically, it is the Kullback-Leibler distance between a bidimensional distribution and the product of its margins (forced independence):

$$I(X,Y) = d_{KL}(p(X,Y)||p(X)p(Y)) = \sum_{x,y} p(x,y) \ln \left(\frac{p(x,y)}{p(x)p(y)} \right).$$

Figure 5: Normalized mutual information by tenure and experience



Notes: Mutual information by tenure and experience across five periods, normalized by the minimum of the entropy. See text for more details.

It thus measures, in our case, the distance between observed and independent matching.[19] A drawback of the standard MI index is that it does not say if the matching is positive or negative assortative (PAM or NAM), but the direction of matching can be easily inferred from a graphical illustration of worker type composition across firm types as shown in Figure 3.

Another drawback of the standard MI is that its range is not bounded (just like the covariance). We shall thus consider a normalized version of the MI (like the correlation). If X is a deterministic function of Y i.e. perfect dependence, then all information conveyed by X is shared with Y. In this case, the mutual information is the same as the uncertainty contained in X (or Y) alone, namely the entropy of X (or Y). We therefore use the following normalized MI,

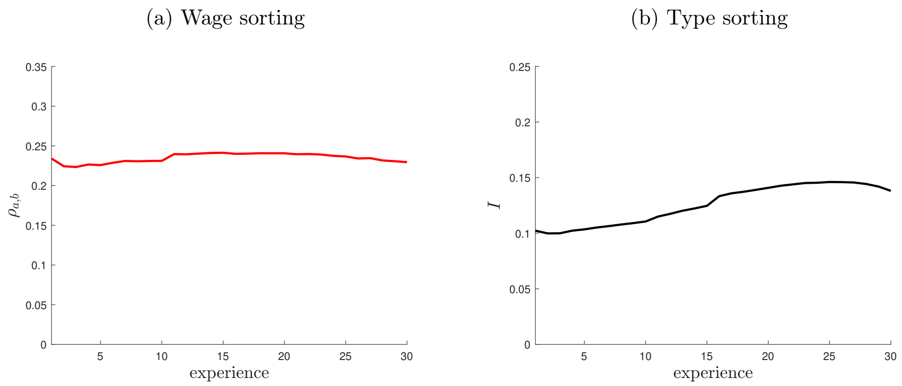
$$\tilde{I}(X,Y) = \frac{I(X,Y)}{\min\left[H(X), H(Y)\right]},$$

where $H(X) = -\Sigma_x p(x)\ln(p(x))$ is the entropy of X, and $H(Y) = -\Sigma_y p(y)\ln(p(y))$ is the entropy of Y. In the extreme case of perfect dependence where X and Y are the same random variable, then $\tilde{I} = 1$, and in the other extreme case of independence, $\tilde{I} = 0$.

Figure 5 shows the normalized MI for the cross-sectional distribution $p(k,\ell,x)$ given

[19]The assortative marriage literature has also measured sorting by comparing the observed and independent matching probabilities between partners. See Greenwood et al., 2016 for example.

Figure 6: Lifecycle sorting



Notes: Panel a shows the correlation between worker and firm wage fixed effect over experience, while panel b displays the mutual information by experience. Both measures are averaged across five periods.

x. The results are starkly different from the wage fixed effect correlations, wherein sorting is increasing in tenure. In Section 5 we emphasized that preference intensity is increasing in tenure, which is consistent with MI being increasing in tenure. Thus, the selection into long tenure jobs (presumably revealed to be preferred) is associated with higher dependence between worker types with particular firm types – resulting in increased type sorting. Significantly, correlations between wage effects do not reveal this sorting, which means that good long-term employment relationships are not necessarily the ones with higher wages. Instead, job preferences may depend upon *non-wage* factors (amenities).

6.2 Type sorting increases with experience

We have already seen that job preferences are more intense when tenure increases, although not when one gets older for a given tenure. However, as a person ages, better opportunities present themselves, employment retention, and subsequent job transitions are less likely to be driven by wages. To better understand the role of mobility on sorting, we simulate cohorts of individuals drawn from the initial distribution. We expect cohorts to become more sorted as they age, but the sorting on wages should be weakening.

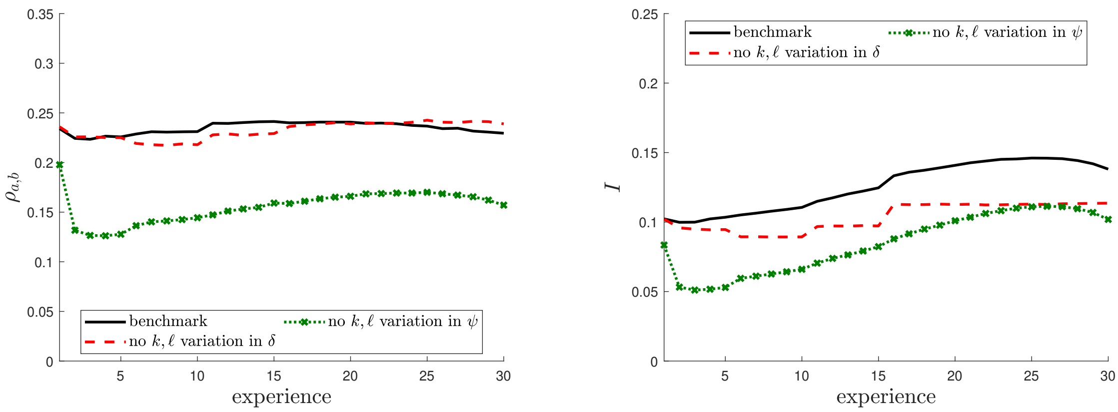
Figure 6 displays the average profiles of sorting measures across the five periods where the profiles of both mutual information and wage fixed effects correlations are qualitatively similar in each period. As expected, the mutual information picks an increasing trend with experience that the wage correlation ignores. Next we investigate what contributes to the rising profile of type sorting.

6.3 Chance matters when young, choice matters when old

The model has four channels of sorting: 1) Job preferences γ; 2) Layoff δ; 3) Market segmentation λ; 4) Reemployment ψ. To gauge the assortative matching importance of

Figure 7: Counterfactual E-U and U-E transitions

(a) Wage sorting (b) Type sorting



Notes: Panel a shows the correlation between worker and firm wage fixed effect over experience, while panel b displays the mutual information by experience. Both measures are averaged across five periods.

a channel over workers' age, we simulate cohorts of individuals drawn from the initial distribution where we counterfactually remove variation in each channel either across k, or across ℓ, or both.[20] Mechanically, the model can arrive at no sorting between worker and firm types if all the channels have either (i) no variation across k, (ii) no variation across ℓ, or (iii) both. Results are qualitatively similar in all three cases; hence, we only present case (iii) in what follows.

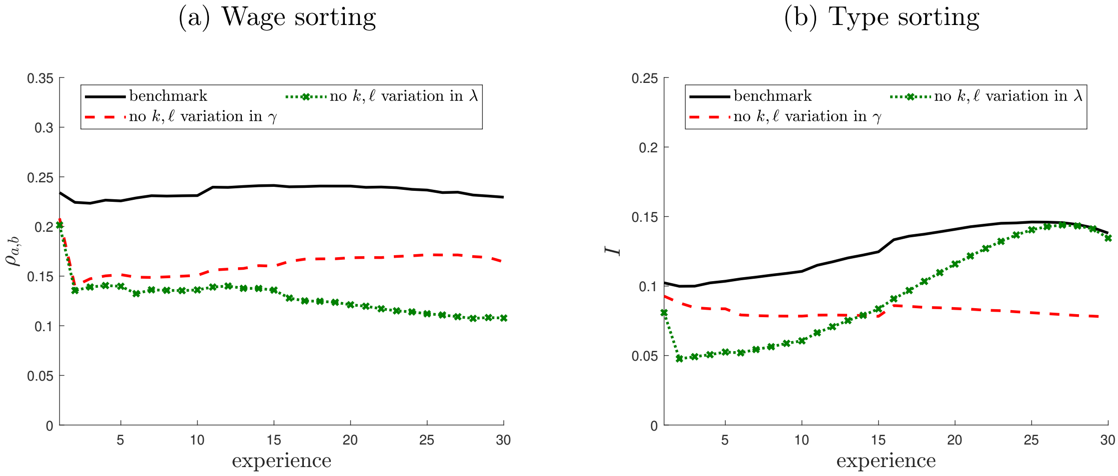
Transitions to and from unemployment. Figure 7 displays the results for the layoff (δ) and reemployment (ψ) channels, for the correlation and the mutual information, respectively. The benchmark simulation is the black, solid line. The counterfactual is to eliminate a given channel, holding all other channels at their estimated values.

Removing heterogeneity from reemployment probabilities ψ results in a significant drop in both sorting indexes. Interestingly, this drop is realized in the first two years of age and then the index starts increasing more or less in parallel to the benchmark index. Removing heterogeneity from layoff probabilities δ has overall no impact on wage sorting, and has an impact on type sorting increasing with age.

In summary, the parameters governing transitions to and from unemployment contribute to overall sorting with the latter being a relatively notable channel. Reemployment probabilities hit sorting strongly when workers are young but less so later on. Layoff hits type sorting more when workers become older.

[20] For example, for the layoff channel we remove k variation by the counterfactual $\hat{\delta}_{k\ell} = \frac{1}{K}\sum_k \delta_{k\ell}$. The preference channel is special in that eliminating ℓ variation cannot be done without also eliminating k variation due to the normalization that $\sum_\ell \gamma_{k\ell} = 1$.

Figure 8: Counterfactual E-E' transitions



Notes: Panel a shows the correlation between worker and firm wage fixed effect over experience, while panel b displays the mutual information by experience. Both measures are averaged across five periods.

Job-to-job transitions. Next, we compare and contrast the relative roles of job preferences γ and market segmentation λ in job-to-job transition probabilities. Figure 8 displays the results, again averaging over five periods where results in each period are qualitatively similar. Job preferences γ plays a significant role in type sorting, measured by the normalized MI index. It is also a stronger source of sorting for more experienced and tenured workers as the cohort ages, reflecting the greater intensity of preferences. Job preferences appear to be the driving force for increased sorting over worker careers. What we see is that sorting naturally tends to increase with age. If we remove the source of sorting incorporated in γ, then the MI indices become flat, and the observed age profile becomes not replicable. The impact on wage correlations is less pronounced.

On the other hand, the effect of λ, which models the offer arrival process, is a lot stronger in the correlation than in the MI index. It suggests a strong contribution to sorting early in careers and reduced importance as workers age.

To sum up, we find that complementarities between worker and firm types in the offer arrival process while employed λ and in the re-employment probabilities ψ are key drivers of sorting during early career. The roles of λ and ψ become apparent in the correlation and thus seem to be driving a classical form of sorting via wage effects. Second, later in life, the way workers rank choices of jobs (γ) dictates their matches and dominates differences in chances to move on-the-job (λ). Moreover, the role of γ is more apparent in the MI index than in the wage fixed effects correlation, and it is what gives rise to the increasing sorting with age. Hence, we conclude that chances are key drivers of sorting when workers are young and choices are key drivers when old. These four channels, however, may interact and drive sorting jointly. We investigate this in the next section.

6.4 The interplay of sorting channels

We end this study by showing how 1) job preferences γ, 2) layoff δ, 3) market segmentation λ and 4) reemployment ψ may interact and jointly affect wage and type sorting. It is natural that in a more restricted mobility model, multiple channels may work in concert. For example, an endogenous job destruction model would view layoff rates to be a function of the preferences for a match and therefore eliminating preference variation would also impact layoff variation.

Specifically, denote by S_{at} the sorting measure (either mutual information or wage effects correlation) for a given simulation at cohort age a and time period t. Denote by $\bar{S} = \sum_{a,t} S_{at}$ the aggregate measure. To measure the relative importance of each channel to overall sorting, we decompose how much each channel contributes to $\bar{S}$ through the sequential elimination of channels. Since the channels interact with each other, the order of elimination matters.[21] For example, consider the channel elimination order 4321, where channel 4 is first eliminated, then 3, 2 and 1. In this case, the marginal contribution of channel 4 is measured by $\left(\bar{S} - \bar{S}_{-4}\right)/\bar{S}$. The marginal contribution of channel 3 is measured by $\left(\bar{S}_{-4} - \bar{S}_{-43}\right)/\bar{S}$, and so forth, where $\bar{S}_{-(\cdot)}$ denotes the aggregate counterfactual measure of sorting. There are 24 different elimination orders. We do channel elimination in 3 ways: no variation across k, no variation across ℓ, and finally no variation across both k and ℓ.[22]

Results are similar across periods and whether we eliminate k-variation, ℓ-variation or both. While the marginal contribution of each channel does vary by age, to focus attention on the overall interactions between the channels, we summarize the detailed picture through an aggregation over age and over the 5 time windows. As before, we present our results from cases where we eliminate both k and ℓ variations. In Table 6, we calculate the importance of each channel considering all possible orders of elimination. We emphasize in bold the relative reduction in sorting of removing each of the four channels separately (direct channel effects).

In line with our previous section, the direct effect of job preferences γ is important (about 30-35%) both in type and wage sorting. The second biggest drop after removing the heterogeneity in job preferences is market segmentation λ (another 40-50%). This is also true when one removes heterogeneity in job offer rates before job preferences. Job preferences and segmentation together explain 75% of sorting both in wages effects and latent types. It thus seems that both components of transition probabilities have independent contributions to sorting (no complementarity). Note that we find the same independence between job preferences γ and reemployment probabilities ψ for type sorting. For wage sorting there is a moderate complementarity that reduces the effect of one channel when it is introduced after the other one.

[21] Taber and Vejlin (2020) face a similar issue.

[22] The preference channel is treated as an exception in that we simply adopt indifference as the elimination of the channel under all three regimes, $\gamma_{k\ell} = 1/L$.

Table 6: Decomposing wage and type sorting

Elimination order	Wage sorting γ	Wage sorting δ	Wage sorting λ	Wage sorting ψ	Type sorting γ	Type sorting δ	Type sorting λ	Type sorting ψ
1234	**0.307**	-0.116	0.467	0.342	**0.350**	-0.090	0.411	0.329
1243	**0.307**	-0.116	0.130	0.680	**0.350**	-0.090	0.248	0.492
1324	**0.307**	-0.133	0.485	0.342	**0.350**	-0.075	0.396	0.329
1342	**0.307**	0.076	0.485	0.132	**0.350**	0.047	0.396	0.206
1423	**0.307**	0.329	0.130	0.235	**0.350**	0.074	0.248	0.328
1432	**0.307**	0.076	0.382	0.235	**0.350**	0.047	0.274	0.328
2134	0.182	**0.009**	0.467	0.342	-0.082	**0.342**	0.411	0.329
2143	0.182	**0.009**	0.130	0.680	-0.082	**0.342**	0.248	0.492
2314	0.380	**0.009**	0.269	0.342	0.225	**0.342**	0.103	0.329
2341	0.092	**0.009**	0.269	0.630	0.044	**0.342**	0.103	0.510
2413	-0.016	**0.009**	0.130	0.878	-0.058	**0.342**	0.248	0.468
2431	0.092	**0.009**	0.021	0.878	0.044	**0.342**	0.145	0.468
3124	0.339	-0.133	**0.452**	0.342	0.492	-0.075	**0.254**	0.329
3142	0.339	0.076	**0.452**	0.132	0.492	0.047	**0.254**	0.206
3214	0.380	-0.174	**0.452**	0.342	0.225	0.191	**0.254**	0.329
3241	0.092	-0.174	**0.452**	0.630	0.044	0.191	**0.254**	0.510
3412	0.449	0.076	**0.452**	0.023	0.675	0.047	**0.254**	0.023
3421	0.092	0.433	**0.452**	0.023	0.044	0.678	**0.254**	0.023
4123	0.202	0.329	0.130	**0.339**	0.351	0.074	0.248	**0.327**
4132	0.202	0.076	0.382	**0.339**	0.351	0.047	0.274	**0.327**
4213	-0.016	0.547	0.130	**0.339**	-0.058	0.483	0.248	**0.327**
4231	0.092	0.547	0.021	**0.339**	0.044	0.483	0.145	**0.327**
4312	0.449	0.076	0.136	**0.339**	0.675	0.047	-0.049	**0.327**
4321	0.092	0.433	0.136	**0.339**	0.044	0.678	-0.049	**0.327**

Notes: Marginal contributions of the four channels in the first column include 1) job preferences γ, 2) layoff δ, 3) market segmentation λ and 4) reemployment ψ. For example, 4123 means removing ψ first, γ second, then δ and λ. The number in the table is the relative change in the correlation or the mutual information index.

The direct effect of layoff rates δ is negligible in wage sorting but sizable (34%) in type sorting. Eliminating the layoff channel implies that job preferences γ stop having any effect on sorting. And after the elimination of the job preference channel, there is no room left for the layoff channel. This indicates a strong complementarity of the job values and layoff channels, particularly for type sorting. We have already shown evidence that the most preferred jobs tend to be those that last longer. If these most preferred jobs stop lasting longer, then they become less prevalent in the population distribution, and further removing job preference heterogeneity matters less.

The second biggest drop after removing layoff heterogeneity (δ) is with reemployment probabilities ψ, and vice versa. Together they explain about 90% of wage sorting and 80% of type sorting. Removing one of these two channels boosts the other one, but a lot more so for wage sorting than for type sorting. We note also that removing the heterogeneity in alternative job offer arrival rates λ affects wage sorting more than type sorting (direct effect of 45% vs 25%). So, it seems that market segmentation, structuring the job offer rates λ as well as the reemployment probabilities ψ, are more important for wage sorting, implying that job offers are more likely to be directed toward matches that increase pay.

Overall, while job preferences γ and market segmentation λ appear to have independent contributions to sorting (no complementarity), this is not the case for the rest of the channels. Layoff rates δ, job preferences γ and reemployment probabilities ψ considerably interact. Our results highlight that sorting is indeed convoluted and multifaceted. Measuring and analyzing sources of sorting require a rich and flexible framework.

7 Conclusion

In this paper we extend the finite mixture framework of Bonhomme et al. (2019) (BLM) to estimate a model of wages and employment mobility with two-sided heterogeneity. We propose a new parameterization for job-to-job transition probabilities, which allows us to meaningfully quantify the relative importance of job arrival process versus job preferences over different stages of workers' careers. However, this parameterization is highly non-linear. We provide a strategy to overcome this estimation challenge by nesting an MM algorithm (Hunter, 2004; Hunter and Lange, 2004) inside the M-step of BLM's EM algorithm. We also nest the whole EM algorithm (with the nested MM algorithm) inside a new firm classification algorithm, aimed at producing a classification of firms based on the full information of the wage-and-mobility likelihood instead of BLM's k-means pre-step.

Our application focuses on Denmark in the period 1989–2013. The results from our research are summarized as follows. Worker type conditional job preferences are revealed through job to job transitions to vary substantially across employer types. We show that the variation has a strong pecuniary component, in particular in terms of the net present value of future earnings associated with the current job. The relationship to current job wages is weaker. The pecuniary motive is particularly pronounced at short tenure and weakens with long tenure. Job preferences intensify both in age, experience, and tenure. We also show that higher wage type workers have more intense preferences.

We introduce a new measure of sorting, the mutual information (MI). We employ MI to accurately represent the dependence between worker and firm types when sorting is convoluted by non-wage factors or non-linearity in wage fixed effects. We find that type sorting, measured by the MI index, increases as workers settle into long tenure relationships. We find the opposite for wage sorting, measured by wage fixed effect correlations. This implies the increased importance of non-wage job characteristics in sorting patterns as workers age and select into long tenure relationships. Consistently, our estimated match preferences, revealed from job-to-job moves, show that as workers age and tenure rises, job preferences are increasingly shaped by non-wage attributes. One such characteristic is employment protection. We find that workers tend to accumulate in firm groups where exit rates and unemployment risk are lower. Moreover, unemployment risk has little effect on wage sorting, whereas it strongly determines type sorting.

In addition, we observe that the offer arrival process while employed and the reemployment probabilities are key drivers of sorting during early career. These two channels are apparent particularly in the wage correlation and thus seem to be driving a classical form of sorting via wage effects. However, as workers age, job preferences dictate their matches and dominate differences in chances to move on the job. The role of job preferences is what gives rise to the positive age trend in type sorting, which materializes more in the MI index than in the wage fixed effects correlation.

Finally, we show that these various channels of sorting interact with each other. Our results highlight that sorting is convoluted and one cannot understand its sources by studying wages alone. There is not a global firm ladder across workers, nor is the ladder purely determined by wages. One general conclusion for future studies on sorting is that it is crucial to use both a flexible estimator taking into account the interaction between wages and mobility as well as a flexible measure of sorting that can accommodate non-linearity and multi-dimensionality of matching. Our extension of BLM and the mutual information index offers a tractable and flexible way to do that, and goes beyond the wage model of AKM.

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., K. L. McKinney, and I. M. Schmutte (2019, July). Modeling Endogenous Mobility in Earnings Determination. *Journal of Business & Economic Statistics 37*(3), 405–418.

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.

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.

Azkarate-Askasua, M. and M. Zerecero (2019). Correcting small sample bias in linear models with many covariates. Technical report, TSE.

Bagger, J. and R. Lentz (2019, 05). An Empirical Model of Wage Dispersion with Sorting. *The Review of Economic Studies 86*(1), 153–190.

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

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy 81*(4), 813–846.

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

Bonhomme, S., T. Lamadon, and E. Manresa (2019, May). A Distributional Framework for Matched Employer Employee Data. *Econometrica 87*(3), 699–739.

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.

Gautier, P., C. N. Teulings, and A. van Vuuren (2010). On-the-job search, mismatch and efficiency. *Review of Economic Studies 77*(1), 245–272.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics 8*(1), 1–41.

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

Guvenen, F. (2009, January). An Empirical Investigation of Labor Income Processes. *Review of Economic Dynamics 12*(1), 58–79.

Hall, R. E. and A. I. Mueller (2018). Wage dispersion and search behavior: The importance of nonwage job values. *Journal of Political Economy 126*(4), 1594–1637.

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.

Hornstein, A., P. Krusell, and G. L. Violante (2011, December). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review 101*(7), 2873–98.

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.

Jarosch, G. (2021). Searching for job security and the consequences of job loss. Technical report.

Kline, P., R. Saggio, and M. Solvsten (2018, June). Leave-out estimation of variance components. Papers 1806.01494, arXiv.org.

Kline, P., R. Saggio, and M. Solvsten (2020, September). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

Lentz, R. (2010). Sorting by search intensity. *Journal of Economic Theory 145*(4), 1436–1452. Search Theory and Applications.

Lentz, R., J. Maibom, and E. Moen (2022). Random or competitive search? *Working Paper*.

Lindenlaub, I. (2017, 01). Sorting Multidimensional Types: Theory and Application. *The Review of Economic Studies 84*(2), 718–789.

Lindenlaub, I. and F. Postel-Vinay (2020). Multidimensional Sorting under Random Search. 2017 Meeting Papers 501.

Lindenlaub, I. and F. Postel-Vinay (2021, January). The Worker-Job Surplus. NBER Working Papers 28402, National Bureau of Economic Research, Inc.

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 F. Postel-Vinay (2020, August). Multidimensional Skills, Sorting, and Human Capital Accumulation. *American Economic Review 110*(8), 2328–2376.

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

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(6), 2295–2350.

Shimer, R. and L. Smith (2000, March). Assortative matching and search. *Econometrica 68*(2), 343–369.

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.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2018, 10). Firming Up Inequality*. *The Quarterly Journal of Economics 134*(1), 1–50.

Sorkin, I. (2018, 01). Ranking Firms Using Revealed Preference. *The Quarterly Journal of Economics 133*(3), 1331–1393.

Taber, C. and R. Vejlin (2020). Estimation of a roy/seach/compensating differential model of the labor market. *Econometrica 88*(3), 1031–1069.

Usui, E. (2008). Job satisfaction and the gender composition of jobs. *Economics Letters 99*, 23–26.

Villanueva, E. (2007, July). Estimating compensating wage differentials using voluntary job changes: Evidence from germany. *Industrial and Labor Relations Review 60*(4).

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

APPENDIX

A Identification

Identification is essentially covered in Bonhomme et al. (2019). We here discuss the argument given our model assumptions, insisting on the main steps.

Firm classification. First, we assume that a vector of firm characteristics (wage moments, size, worker entry/exit statistics, etc) exists that provides enough information to classify firms into L groups $\ell \in \{1, ..., L\}$.

Job-to-job moves. Second, consider all possible worker trajectories with 3 wages (w_1, w_2, w_3), where the first wage corresponds to a job of type ℓ_1 and the last couple to a job of type ℓ_2, that is there is a job mobility between periods 1 and 2. The likelihood of this event (conditional on ℓ_1-employment) is

$$p_{\ell_1\ell_2}(w_1, w_2, w_3) = \sum_k \pi_{\ell_1\ell_2}(k) f_{\ell_1}(w_1|k) g_{\ell_2}(w_2, w_3|k) \tag{10}$$

where g is the joint distribution of (w_2, w_3) for a short tenure worker and f is the cross-sectional distribution of any wage and π is the overall employment probability:

$$\pi_{\ell_1\ell_2}(k) = M_{\ell_1\ell_2}(k) M_{\ell_1}^{-}(k). \tag{11}$$

Assuming that $M_{\ell_1\ell_2}(k) > 0$ and $\sum_{\ell_2} M_{\ell_1\ell_2}(k) < 1$, all worker groups are represented in all likelihoods. Moreover, given that $M_{\ell_1}^{-}(k) = 1 - \sum_{\ell_2} M_{\ell_1\ell_2}(k)$, the probability of not moving is a function of $M_{\ell_1\ell_2}(k)$, we can solve for $M_{\ell_1\ell_2}(k)$ given $\pi_{\ell_1\ell_2}(k)$ (quadratic equation).

The idea is to store these probabilities in a matrix

$$P_{\ell_1\ell_2} = [p_{\ell_1\ell_2}(w_1, w_2, w_3)]_{w_1,(w_2,w_3)}$$

where we index rows by the values of w_1 and the columns by the values of (w_2, w_3). This is assuming discrete wages. If wages are continuous, we can work with CDFs and a wage grid as in BLM. This argument is essentially the same.

It follows from equation (10) that

$$P_{\ell_1\ell_2} = F_{\ell_1} D_{\ell_1\ell_2} G_{\ell_2}^{\top}$$

where

$$F_{\ell_1} = [f_{\ell_1}(w_1)]_{w_1,k}\,, D_{\ell_1\ell_2} = \operatorname{diag}[\pi_{\ell_1\ell_2}(k)]_k\,, G_{\ell_2} = [g_{\ell_2}(w_2, w_3)]_{(w_2,w_3),k}\,.$$

BLM use a more sophisticated argument involving general mobility cycles. Let us simply assume that all transitions (ℓ_1, ℓ_2) have positive probability for any worker type. We also assume that the matrices F_ℓ, G_ℓ are full column rank and that the diagonal matrices D_ℓ have all their K diagonal entries non nil.

Fix $\ell_1 = \ell$. There exists a singular value decomposition: $P_{\ell\ell} = U_\ell \Lambda_\ell V_\ell^\top$, where U_ℓ and V_ℓ are orthogonal matrices with $U_\ell^\top U_\ell = I_N$, $V_\ell^\top V_\ell = I_M$ (say $M = 2N$) and Λ_ℓ is a rectangular diagonal matrix containing the singular values. The number of non-zero diagonal entries in Λ_ℓ is equal to the number of groups K. Let $\Lambda_{\ell 11}$ be the (K, K) diagonal matrix containing the non-zero singular values, and let $U_\ell = (U_{\ell 1}, U_{\ell 2})$ and $V_\ell = (V_{\ell 1}, V_{\ell 2})$ partition the columns of U_ℓ, V_ℓ so that $P_{\ell\ell} = U_{\ell 1}\Lambda_{\ell 11} V_{\ell 1}^\top$.

Note also that since the columns of U_ℓ are orthogonal vectors,

$$U_{\ell 2}^\top P_{\ell\ell} = U_{\ell 2}^\top U_{\ell 1}\Lambda_{\ell 11} V_{\ell 1}^\top = 0_{(N-K)\times M}.$$

Hence,

$$U_{\ell 2}^\top P_{\ell\ell} = U_{2\ell}^\top F_\ell D_{\ell\ell} G_\ell^\top = 0_{(N-K)\times M}.$$

As $D_{\ell\ell} G_\ell^\top$ is a full row-rank (K, M) matrix, it follows that $U_{\ell 2}^\top F_\ell = 0_{(N-K)\times K}$. Symmetrically, $P_{\ell\ell} V_{\ell 2} = 0$ since $V_{\ell 1}^\top V_{\ell 2} = 0$. Now, since $F_\ell D_{\ell\ell}$ has rank K, it follows that $G_\ell^\top V_{\ell 2} = 0_{K\times(M-K)}$.

Next, we deduce from $P_{\ell\ell} = U_{\ell 1}\Lambda_{\ell 11} V_{\ell 1}^\top$ that

$$\Lambda_{\ell 11}^{-1} U_{\ell 1}^\top F_\ell D_{\ell\ell} G_\ell^\top V_{\ell 1} = I_K.$$

Given this result, if we define $W_\ell = \Lambda_{\ell 11}^{-1} U_{\ell 1}^\top F_\ell$, then, $W_\ell^{-1} = D_{\ell\ell} G_\ell^\top V_{\ell 1}$.

It is important to note that all these W_ℓ matrices have been calculated independently for all ℓ. Their columns correspond to different k's but there is no way, yet, to make sure that the columns of W_ℓ for different ℓ's are consistently labeled.

Consider any ℓ_2. If W_ℓ and W_{ℓ_2} have been constructed with a compatible ordering of rows and columns, then

$$Q_{\ell\ell_2} := \Lambda_{\ell 11}^{-1} U_{\ell 1}^\top P_{\ell\ell_2} V_{\ell_2 1} = \Lambda_{\ell 11}^{-1} U_{\ell 1}^\top F_\ell D_{\ell\ell_2} G_{\ell_2}^\top V_{\ell_2 1} = W_\ell D_{\ell\ell_2} D_{\ell_2\ell_2}^{-1} W_{\ell_2}^{-1}.$$

Therefore, the first thing to do is to find the permutation of the rows and columns of W_{ℓ_2}, and of the columns of $U_{\ell_2 1}$ and $V_{\ell_2 1}$, that makes $W_\ell^{-1} Q_{\ell\ell_2} W_{\ell_2}$diagonal.

Next, consider $Q_{\ell\ell_2}$ and $Q_{\ell_2\ell}$. We have

$$Q_{\ell\ell_2} Q_{\ell_2\ell} = W_\ell D_{\ell\ell_2} D_{\ell_2\ell_2}^{-1} D_{\ell_2\ell} D_{\ell\ell}^{-1} W_\ell^{-1}.$$

Assuming that the diagonal entries of $D_{\ell\ell_2} D_{\ell_2\ell_2}^{-1} D_{\ell_2\ell} D_{\ell\ell}^{-1}$ are all distinct, they are thus identifiable as the eigenvalues of $Q_{\ell\ell_2} Q_{\ell_2\ell}$ and the eigenvectors are identified up to a multiplicative constant. Let $\widehat{W}_\ell$ be the matrix of orthonormal eigenvectors, then

$$\widehat{W}_\ell = W_\ell \Delta_\ell = \Lambda_{\ell 11}^{-1} U_{\ell 1}^\top F_\ell \Delta_\ell$$

for some diagonal matrix Δ_ℓ.

We cannot immediately deduce that $U_{\ell 1}\Lambda_{\ell 11}\widehat{W}_\ell = U_{\ell 1}U_{\ell 1}^\top F_\ell \Delta_\ell = F_\ell \Delta_\ell$ because $U_\ell U_\ell^\top = I_N$ does not imply that $U_{\ell 1}U_{\ell 1}^\top = I_N$. Still, $F_\ell \Delta_\ell = U_{\ell 1}\Lambda_{\ell 11}\widehat{W}_\ell$ because, as $U_{\ell 2}^\top F_{\ell_2}\Delta = 0_{(N-K)\times K}$, we also have

$$\begin{pmatrix} \Lambda_{\ell 11}\widehat{W}_\ell \\ 0_{(N-K)\times K} \end{pmatrix} = U_\ell^\top F_\ell \Delta_\ell.$$

Hence,

$$U_{\ell 1}\Lambda_{\ell 11}\widehat{W}_\ell = U_\ell \begin{pmatrix} \Lambda_{\ell 11}\widehat{W}_\ell \\ 0_{(N-K)\times K} \end{pmatrix} = U_\ell U_\ell^\top F_\ell \Delta = F_\ell \Delta_\ell.$$

From $F_\ell \Delta_\ell = U_{\ell 1}\Lambda_{\ell 11}\widehat{W}_\ell$ we deduce $F_\ell = U_{\ell 1}\Lambda_{\ell 11}\widehat{W}_\ell \Delta_\ell^{-1}$. Since the rows of F_ℓ sum to one (each column being a probability distribution), then Δ_ℓ is identified (since all its diagonal terms Δ_k are identified). Hence, F_ℓ is identified.

Finally, knowing W_ℓ we know $W_\ell^{-1} = D_{\ell\ell} G_\ell^\top V_{\ell 1}$. A similar argument allows to separately identify the diagonal matrix $D_{\ell\ell}$ and G_ℓ. (The rows of G_ℓ sum to one.)

Long tenure, no mobility. Consider the continuation of the preceding trajectories in the same firm ℓ_2 with two additional wages (w_4, w_5). Call this new spell long tenure. The likelihood is

$$p_{\ell_1\ell_2}(w_1, w_2, w_3, w_4, w_5) = \sum_k \widetilde{\pi}_{\ell_1\ell_2}(k) f_{\ell_1}(w_1|k) g_{\ell_2}^{ST}(w_2, w_3|k) g_{\ell_2}^{LT}(w_4|k, w_3)$$

where the transition probability $\widetilde{\pi}$ incorporates the additional no-move event. This likelihood easily identifies $g_{\ell_2}^{LT}(w_4|k, w_3)$.

B An MM algorithm for the M-step update of wage densities and transition probabilities

In the M-step of the standard EM algorithm, model parameters are updated to maximize the expected log likelihood of the data, which up to a constant is a minorizing function of the actual likelihood of the data, see for example Hunter and Lange (2004). In so doing, the M step finds parameters that necessarily represent a likelihood function improvement relative to the minorizing function's defining point. Unlike the standard EM algorithm applications, our expected log likelihood is not linear in some of our parameters (specifically, the ones related to the mobility model). Thus, we modify the expected log likelihood to obtain a minorization of the data likelihood that is indeed linear in our parameters and therefore allow an easy maximization of the minorizing function. In this appendix we document the first order conditions associated with the maximization of the minorizing function as well as the modifications of it relative to the standard expected log likelihood.

B.1 Wage parameters

The first wage in a job spell is drawn from the static wage distribution,

$$f_{static}(w|k,\ell,x) = \frac{1}{\omega_{k\ell}(x)}\varphi\left(\frac{w-\mu_{k\ell}(x)}{\omega_{k\ell}(x)}\right),$$

and subsequent wages within a job evolve according to the density,

$$f_{dyn}(w'|k,\ell',x',w,\ell,x) = \frac{1}{\sigma_{k\ell}(x')}\varphi\left(\frac{w'-\mu_{k\ell'}(x')-\rho\left[w-\mu_{k\ell}(x)\right]}{\sigma_{k\ell}(x')}\right),$$

where (w',ℓ',x') refer one period forward relative to (w,ℓ,x).

The wage part of the expected log likelihood is given by,

$$\begin{aligned} W = \sum_{i,k} p_i\left(k|\beta,F\right)\sum_{t=1}^{T_i}\Big[& D_{it-1}\ln f_{static}(w_{it}|k,\ell_{it},x_{it})+ \\ & (1-D_{it-1})\ln f_{dyn}\left(w_{it}|k,\ell_{it},x_{it},w_{it-1},\ell_{it-1},x_{it-1}\right)\Big], \end{aligned}$$

where for convenience, $D_{i0}\equiv 1$. While the types (k,ℓ) remain the same within a job, wage parameters still vary across experience and tenure status. Let $\overline{\mu}$ denote the mean wage parameters of (k,ℓ,x) cell that we are trying to estimate.

$$\begin{aligned} \frac{dW}{d\overline{\mu}} = & \sum_{i,k} p_i(k|\beta,F)\sum_{t=1}^{T_i}\frac{1}{\sigma_{k\ell}(x_{it})}\Bigg[D_{it-1}\frac{d\mu_{k\ell_t}\left(x_{it}\right)}{d\overline{\mu}}\left(w_{it}-\mu_{k\ell_t}\left(x_{it}\right)\right)+ \\ & (1-D_{it-1})\left(\frac{d\mu_{k\ell_t}\left(x_{it}\right)}{d\overline{\mu}}-\rho\frac{d\mu_{k\ell_{t-1}}\left(x_{it-1}\right)}{d\overline{\mu}}\right)\left(w_{it}-\mu_{k\ell_t}\left(x_{it}\right)-\rho\left(w_{it-1}-\mu_{k\ell_{t-1}}\left(x_{it-1}\right)\right)\right)\Bigg] = 0. \end{aligned}$$

For $D_{it-1}=0$, there are three cases for the summand,

$$\begin{aligned} S_{ikt} = & D_{it-1}\frac{d\mu_{k\ell_t}\left(x_{it}\right)}{d\overline{\mu}}\left[w_{it}-\mu_{k\ell_t}\left(x_{it}\right)\right]+ \\ & (1-D_{it-1})\left[\frac{d\mu_{k\ell_t}\left(x_{it}\right)}{d\overline{\mu}}-\rho\frac{d\mu_{k\ell_{t-1}}\left(x_{it-1}\right)}{d\overline{\mu}}\right]\left[w_{it}-\mu_{k\ell_t}\left(x_{it}\right)-\rho\left(w_{it-1}-\mu_{k\ell_{t-1}}\left(x_{it-1}\right)\right)\right], \end{aligned}$$

1. $\frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}}=0, \frac{d\mu_{k\ell_{t-1}}(x_{it-1})}{d\overline{\mu}}=1$ then $S_{ikt}=-\rho\left[w_{it}-\mu_{k\ell_t}\left(x_{it}\right)-\rho\left(w_{it-1}-\overline{\mu}\right)\right].$
2. $\frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}}=1, \frac{d\mu_{k\ell_{t-1}}(x_{it-1})}{d\overline{\mu}}=0$ then $S_{ikt}=w_{it}-\overline{\mu}-\rho\left(w_{it-1}-\mu_{k\ell_{t-1}}\left(x_{it-1}\right)\right).$
3. $\frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}}=1, \frac{d\mu_{k\ell_{t-1}}(x_{it-1})}{d\overline{\mu}}=1$ then $S_{ikt}=\left[1-\rho\right]\left[w_{it}-\rho w_{it-1}-(1-\rho)\overline{\mu}\right]=0.$

We solve for a set of $\mu_{k\ell}(x)$ for a given k using the following procedure. For a given k, let G denote the total possible combinations of ℓ and x, and let g index the order of this combination. Construct A, a square matrix $G \times G$ with each element in row g, column g' containing $A_{gg'}$ where

$$\begin{aligned} A_{gg'} = & \sum_{i,k} p_i(k|\beta, F) \sum_{t=1}^{T_i} \frac{1}{\sigma_{k\ell}(x_{it})} \Bigg[D_{it-1} \frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}_g} \frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}_{g'}} + \\ & (1 - D_{it-1}) \left(\frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}_g} - \rho \frac{d\mu_{k\ell_{t-1}}(x_{it-1})}{d\overline{\mu}_g} \right) \left(\frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}_{g'}} - \rho \frac{d\mu_{k\ell_{t-1}}(x_{it-1})}{d\overline{\mu}_{g'}} \right) \Bigg]. \end{aligned}$$

Let b be a column vector $1 \times G$ with each element

$$b_g = \sum_{i,k} p_i(k|\beta, F) \sum_{t=1}^{T_i} \frac{1}{\sigma_{k\ell}(x_{it})} \left[D_{it-1} w_{it} + (1 - D_{it-1}) \left(\frac{d\mu_{k\ell_t}(x_{it})}{d\overline{\mu}_g} - \rho \frac{d\mu_{k\ell_{t-1}}(x_{it-1})}{d\overline{\mu}_g} \right) [w_{it} - \rho w_{it-1}] \right]$$

We have

$$A \begin{pmatrix} \mu_1 \\ : \\ : \\ \mu_G \end{pmatrix} = b,$$

which by inversion of A delivers $\mu_{k\ell}(x)$ for a given k.

The first order conditions with respect to σ, ω, and ρ deliver,

$$\begin{aligned} \sigma^2_{k\ell}(x) &= \frac{\sum_{i=1} p_i(k|\beta, F) \sum_{t=1} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x, D_{it-1} = 0\} \left[w_{it} - \mu_{k\ell_t}(x_{it}) - \rho \left(w_{it-1} - \mu_{k\ell_{t-1}}(x_{it-1}) \right) \right]}{\sum_{i=1} p_i(k|\beta, F) \sum_{t=1} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x, D_{it-1} = 0\}} \\ \omega^2_{k\ell}(x) &= \frac{\sum_{i=1} p_i(k|\beta, F) \sum_{t=1} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x, D_{it-1} = 1\} [w_{it} - \mu_{k\ell_t}(x_{it})]^2}{\sum_{i=1} p_i(k|\beta, F) \sum_{t=1} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x, D_{it-1} = 1\}} \\ \rho &= \frac{\sum_{i=1} p_i(k|\beta, F) \sum_{t=1}^{T_i} \frac{1 - D_{it-1}}{\sigma_{k\ell}(x_{it})} (w_{it} - \mu_{k\ell_t}(x_{it})) \left(w_{it-1} - \mu_{k\ell_{t-1}}(x_{it-1}) \right)}{\sum_{i=1} p_i(k; |\beta, F) \sum_{t=1}^{T_i} \frac{1 - D_{it-1}}{\sigma_{k\ell}(x_{it})} \left(w_{it-1} - \mu_{k\ell_{t-1}}(x_{it-1}) \right)^2} \end{aligned}$$

B.2 Mobility parameters

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 parameters are

$$
\begin{aligned}
M_{k0\neg} &= 1 - \sum_{\ell'=1}^{L} M_{k0\ell'}, \quad M_{k0\ell'} = \psi_{k\ell'}, \\
M_{k\ell\neg} &= 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}, \quad \ell \geq 1, \\
M_{k\ell\ell'} &= \lambda_{k\ell'} P_{k\ell\ell'}, \quad P_{k\ell\ell'} = \frac{\gamma_{k\ell'}}{\gamma_{k\ell} + \gamma_{k\ell'}}, \quad \ell, \ell' \geq 1, \\
M_{k\ell 0} &= \delta_{k\ell}, \quad \ell \geq 1,
\end{aligned}
$$

and data are

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

where $\#\{\}$ denotes the cardinality of a set and where we reintroduce the control $x_{it} = x$ to remind that we are estimating different parameters for all different control values x. In the rest of this section, we omit the reference to controls x and to $\beta^{(m)}$.

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

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

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 $\ln M_{k\ell\neg}$. An MM algorithm can still be developed as follows.[23]

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

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

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

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

With obvious notations, for $\ell = 1, ..., L$, we can write

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

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

$$\begin{aligned} \ln M_{k\ell\neg} &= \ln \left(1 - \delta_{k\ell} - \sum_{\ell'=1}^{L} \lambda_{k\ell'} + \sum_{\ell'=1}^{L} \lambda_{k\ell'}(1 - P_{k\ell\ell'}) \right) \\ &\geq \ln \underline{M}_{k\ell\neg} \equiv \frac{1 - \delta_{k\ell}^{(s)} - \sum_{\ell'=1}^{L} \lambda_{k\ell'}^{(s)}}{M_{k\ell\neg}^{(s)}} \ln \left(\frac{1 - \delta_{k\ell} - \sum_{\ell'=1}^{L} \lambda_{k\ell'}}{1 - \delta_{k\ell}^{(s)} - \sum_{\ell'=1}^{L} \lambda_{k\ell'}^{(s)}} M_{k\ell\neg}^{(s)} \right) \\ &\qquad + \sum_{\ell'=1}^{L} \frac{\lambda_{k\ell'}^{(s)}(1 - P_{k\ell\ell'}^{(s)})}{M_{k\ell\neg}^{(s)}} \ln \left(\frac{\lambda_{k\ell'}(1 - P_{k\ell\ell'})}{\lambda_{k\ell'}^{(s)}(1 - P_{k\ell\ell'}^{(s)})} M_{k\ell\neg}^{(s)} \right). \end{aligned} \tag{12}$$

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

$$\sum_{k=1}^{K} \sum_{\ell=0}^{L} \left\{ n_{k\ell\neg} \ln \underline{M}_{k\ell\neg} + \sum_{\ell'=0}^{L} n_{k\ell\ell'} \ln M_{k\ell\ell'} \right\}$$

instead of the initial objective.

Let

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

This is the predicted fraction of stayers such as home beats visitor ℓ'.

One can update $\gamma^{(s)}$ so as to maximize

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

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

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

with equality when $\gamma = \gamma^{(s)}$ (see Hunter, 2004), we can instead maximize

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

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

$$\gamma_{k\ell}^{(s+1)} \propto \sum_{\ell'=1}^{L}(\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell'\ell}) \Bigg/ \sum_{\ell'=1}^{L}\frac{\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'} + \widetilde{n}_{k\ell'\ell}^{(s)} + n_{k\ell'\ell}}{\gamma_{k\ell}^{(s)}+\gamma_{k\ell'}^{(s)}},$$

where the proportionality symbol means that the $\gamma_{k\ell}^{(s+1)}$'s should add up to one.

One can update $\delta_{k\ell}^{(s)}, \lambda_{k\ell'}^{(s)}$ by maximizing

$$\sum_{\ell=1}^{L}A_{k\ell}\ln\left(1-\delta_{k\ell}-\sum_{\ell'=1}^{L}\lambda_{k\ell'}\right)+\sum_{\ell=1}^{L}n_{k\ell0}\ln\delta_{k\ell}+\sum_{\ell'=1}^{L}\left(B_{k\ell'}\ln\lambda_{k\ell'}\right),$$

where

$$A_{k\ell} = n_{k\ell\neg}\frac{1-\delta_{k\ell}^{(s)}-\sum_{\ell'=1}^{L}\lambda_{k\ell'}^{(s)}}{M_{k\ell\neg}^{(s)}},\quad B_{k\ell'} = \sum_{\ell=1}^{L}(\widetilde{n}_{k\ell\ell'}^{(s)} + n_{k\ell\ell'}).$$

The FOC for $\delta_{k\ell}$ is

$$-\frac{A_{k\ell}}{1-\delta_{k\ell}-\sum_{\ell'=1}^{L}\lambda_{k\ell'}}+\frac{n_{k\ell0}}{\delta_{k\ell}} = 0 \Rightarrow \delta_{k\ell} = \frac{n_{k\ell0}}{A_{k\ell}+n_{k\ell0}}\left(1-\sum_{\ell'=1}^{L}\lambda_{k\ell'}\right)$$

and

$$1-\delta_{k\ell}-\sum_{\ell'=1}^{L}\lambda_{\ell'} = \frac{A_{k\ell}}{A_{k\ell}+n_{k\ell0}}\left(1-\sum_{\ell'=1}^{L}\lambda_{k\ell'}\right).$$

The FOC for $\lambda_{k\ell'}$ is

$$-\sum_{\ell=1}^{L}\frac{A_{k\ell}}{1-\delta_{k\ell}-\sum_{\ell'=1}^{L}\lambda_{k\ell'}}+\frac{B_{k\ell'}}{\lambda_{k\ell'}} = 0 \Leftrightarrow -\frac{\sum_{\ell=1}^{L}\left(A_{k\ell}+n_{k\ell0}\right)}{1-\sum_{\ell'=1}^{L}\lambda_{k\ell'}}+\frac{B_{k\ell'}}{\lambda_{k\ell'}} = 0,$$

after substituting out $\delta_{k\ell}$.

Let

$$C_{k\ell'} = \frac{B_{k\ell'}}{\sum_{\ell=1}^{L}\left(A_{k\ell}+n_{k\ell0}\right)},\quad \overline{C}_k = \sum_{\ell'=1}^{L}C_{k\ell'}.$$

We finally obtain the following updating formulas:

$$1-\sum_{\ell'=1}^{L}\lambda_{k\ell'}^{(s+1)} = \frac{1}{1+\overline{C}_k} = \frac{\sum_{\ell=1}^{L}\left(A_{k\ell}+n_{k\ell0}\right)}{\sum_{\ell=1}^{L}\left(A_{k\ell}+n_{k\ell0}\right)+\sum_{\ell'=1}^{L}B_{k\ell'}},$$

$$\lambda_{k\ell'}^{(s+1)} = \frac{C_{k\ell'}}{1+\overline{C}_k} = \frac{B_{k\ell'}}{\sum_{\ell=1}^{L}\left(A_{k\ell}+n_{k\ell0}\right)+\sum_{\ell'=1}^{L}B_{k\ell'}}$$

and

$$\delta_{k\ell}^{(s+1)} = \frac{n_{k\ell 0}}{A_{k\ell} + n_{k\ell 0}} \frac{1}{1 + \overline{C}_k} = \frac{n_{k\ell 0}}{A_{k\ell} + n_{k\ell 0}} \frac{\sum_{\ell=1}^{L} (A_{k\ell} + n_{k\ell 0})}{\sum_{\ell=1}^{L} (A_{k\ell} + n_{k\ell 0}) + \sum_{\ell'=1}^{L} B_{k\ell'}}.$$

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

C Numerical implementation (online appendix)

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.

C.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 unemployment).

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 unemployment 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.

C.2 E-step

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

Each CPU holds its own copy of the firm classification, $\mathcal{L}$. With this, CPU c evaluates $L_i(\beta, \mathcal{L}) = \sum_{k=1}^{K} L_i(k; \beta, \mathcal{L})$ for any $i \in \iota_c$ by walking through the worker i linked list of spells. CPU c calculates $L^c = \sum_{i \in \iota_c} \ln L_i(\beta, \mathcal{L})$. The likelihood of the data is then found by summing L^c across CPUs, $L(\beta, \mathcal{L}) = \exp\left(\sum_{c=1}^{C} L^c\right)$. 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.

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

CPU c updates worker posteriors for all $i \in \iota_c$ by, $p_i(k; \beta, \mathcal{L}) = L_k(k; \beta, \mathcal{L})/L_i(\beta, \mathcal{L})$. 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.

C.3 M step

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

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

With the worker posteriors in hand CPU c calculates $\pi_{k,c}(z) = \sum_{i \in \iota_c} p_i(k; \beta, \mathcal{L}) \mathbf{1}\{z_i = z\}$, which is communicated across the CPUs. This is a $K \times Z$ dimension double precision array communication across C CPUs where each CPU receives $\sum_{c=1}^{C} \pi_{k,c}(z)$.[24] Each CPU then calculates $\pi_k(z) = \sum_{c=1}^{C} \pi_{k,c}(z) / \left[\sum_{k=1}^{K} \sum_{c=1}^{C} \pi_{k,c}(z)\right]$.

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

CPU c calculates $m_{k\ell,c}(x) = \sum_{i \in \iota_c} p_i(k; \beta, \mathcal{L}) \mathbf{1}\{x_{i1} = x, \ell_{i1} = \ell\}$, 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 $m_{k\ell}(x) = \sum_{c=1}^{C} m_{k\ell,c}(x) / \left[\sum_{\ell=1}^{L} \sum_{c=1}^{C} m_{k\ell,c}(x)\right]$.

C.3.3 Wage parameters for given $(\beta, \mathcal{L})$.

CPU c calculates

$$\mu_{k\ell,c}(x) = \sum_{i \in \iota_c} p_i(k; \beta, \mathcal{L}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\} w_{it}$$

$$d_{k\ell,c}(x) = \sum_{i \in \iota_c} p_i(k; \beta, \mathcal{L}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}.$$

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 $\sigma_{k\ell,c}(x) = \sum_{i=1}^{I} p_i(k; \beta, \mathcal{L}) \sum_{t=1}^{T} \mathbf{1}\{\ell_{it} = \ell, x_{it} = x\}[w_{it} - \mu_{k\ell}(x)]^2$. 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 $\sigma_{k\ell}(x) = \sqrt{\sum_{c=1}^{C} \sigma_{k\ell,c}(x) / \sum_{c=1}^{C} d_{k\ell,c}(x)}$.

C.3.4 Mobility parameters for given $(\beta, \mathcal{L})$.

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

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

and

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

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 B.

[24] Using mpi_allreduce.

C.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,

$$\ell_j^{(s+1)} = \arg\max_{\ell} \sum_{i=1}^{I} \sum_{k=1}^{K} p_i(k; \widehat{\beta}^{(s)}, \mathcal{L}^{(s)}) \ln L_i(k; \widehat{\beta}^{(s)}, \mathcal{L}_{-j}^{(s)}(\ell)), \tag{13}$$

where $\mathcal{L}_{-j}^{(s)}(\ell)$ is the firm classification that is obtained by taking the $\mathcal{L}^{(s)}$ 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 $q(\ell, \mathcal{L}_{-j}^{(s)}(\ell))$ distribution. The latter does involve all spells but in a way that allows simplification. Define by $\Omega(\mathcal{L})$, the contribution to the expected log likelihood from the $q(\cdot|\mathcal{L})$ related terms,

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

Define,

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

with which we can write,

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

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,

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

where the number of spells in type ℓ firms is,

$$\hat{n}(\ell|\mathcal{L}) \equiv \sum_{j=1}^{J} \hat{n}_j \mathbf{1}\left[\ell_j = \ell\right]. \tag{14}$$

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,

$$\ell_j^{(s+1)} = \arg\max_{\ell} \Bigg[\sum_{(i,t)\in\iota(j)} \sum_{k=1}^{K} p_i(k; \widehat{\beta}^{(s)}, \mathcal{L}^{(s)}) \times \Big[f_{k\ell}(w_{it}|x_{it}) + (1 - D_{it}) \ln \overline{M}_{k\ell_{it}}(x_{it}) + D_{i(t-1)} \ln M_{k\ell_{i(t-1)}\ell} + D_{it} \ln M_{k\ell\ell_{i(t+1)}} \Big] + \Omega(\beta, \mathcal{L}_{-j}^{(s)}(\ell)) \Bigg]. \tag{15}$$

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 $\mathcal{L}^{(s)}$. Denote by $\mathcal{L}^{(s),0} = \mathcal{L}^{(s)}$, where $\mathcal{L}^{(s),j}$ is the firm classification in the jth substep of the C-step. Initialize the C-step by the determination of $\hat{n}(\ell|\mathcal{L}^{(s)})$ by equation (14).

3. Take firm $j = 1$. Find the optimal firm type for firm j according to equation (15) and firm classification $\mathcal{L}^{(s),j-1}$. The (i,t) pairs in $\iota(j)$ are by the data delegation spread out across different CPUs. Each CPU evaluates the summation in equation

(15) 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 (15). 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 (15), and communicates the optimal firm type $\ell_j^{(s+1)}$ to all CPUs, a single integer.

4. Update the firm classification $\mathcal{L}^{(s),j} = \mathcal{L}_{-j}^{(s),(j-1)}(\ell_j^{(s+1)})$. 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 $\hat{n}^j(\ell) = \hat{n}(\ell|\mathcal{L}^{(s),j})$ and the type frequencies $q(\ell|\mathcal{L}^{(s),j})$. 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 $\mathcal{L}^{(s+1)} = \mathcal{L}^{(s),J}$.

D Variance decomposition

As discussed in section 2.4, both stayers and movers share the same mean wage $\mu_{k\ell}$ in our model; this allows us to perform a variance decomposition as in the AKM literature. We decompose the log-wage variance following the projection in (9). Specifically, we expand the initial sample so that each individual observation i is repeated K times, one for each k with an associated weight equal to the estimated posterior probability $p_i(k)$. Then, we decompose the cross-sectional log-wage variance into the effects of experience and tenure ($\overline{\mu}$), worker types (a_k), firm types (b_ℓ), match effects ($\widetilde{\mu}_{k\ell} = \mu_{k\ell} - a_k - b_\ell$), and the residual ($w - \mu_{k\ell}$).

Table 7 shows the log-wage variance decomposition over time. The between (x,k,ℓ)-groups and the within-group variances have rather similar contributions, although when the economy is depressed (as in periods 1 and 5), the residual variance contributes much less (up to 50% and down to 40%). The main between-variance contributor is the worker effect a_k, at just around 25%. After the residual and the worker effect comes the match effect $\widetilde{\mu}_{k\ell}$ that measures the degree of nonlinear interaction between worker and firm effects, whose contribution is decreasing over time (from 15% to 10%). Next, we have the firm effect b_ℓ (from 8% to 5%) and the sorting effect $2\mathrm{cov}(a,b)$ (from 5% to 7%). We thus confirm an increasing sorting trend that has been previously observed (Bagger et al., 2013; Card et al., 2013; Song et al., 2015). In terms of the wage correlation, the increasing sorting trend is monotone over our quarter-century study period with an initial correlation

Table 7: Log-wage variance decomposition (percents of total variance) and wage correlation

	89-93	94-98	99-03	04-08	09-13
residual	42.7	46.4	49.0	50.8	39.6
worker effect, a_k	23.6	22.7	19.3	22.8	25.8
match effect, $\widetilde{\mu}_{k\ell}(x)$	15.6	14.2	13.1	11.5	10.4
firm effect, b_ℓ	8.3	7.0	8.8	5.0	5.4
sorting effect, $2\text{cov}(a_k, b_\ell)$	5.4	5.0	6.1	5.3	6.8
tenure and experience, x	4.4	4.8	4.2	4.6	11.9
$\overline{\mu}(x)$	1.3	2.4	1.5	2.5	4.4
$2\text{cov}(a_k, \overline{\mu}(x))$	2.6	2.0	1.9	1.7	5.9
$2\text{cov}(\overline{\mu}(x), b_\ell)$	0.5	0.4	0.3	0.4	1.6
total between	57.3	53.6	51.0	49.2	60.4
wage correlation, $\text{corr}(a_k, b_\ell)$	0.19	0.20	0.23	0.25	0.29

of about 0.2 increasing to about 0.3. Finally and interestingly, the contribution of tenure and experience $\overline{\mu}(x)$, which used to be small (around 4% in total, including covariances), has increased in the last period. We lack data to decide whether this is significant or not.

We thus find that the contribution of worker heterogeneity is considerably less than the usual AKM estimates reported in Table 8, and the residual variance is considerably higher. Furthermore, the correlation between worker and firm effects in this literature is all over the place—sometimes negative, often zero, and sometimes positive, in any case seldom as large as .2 or .3. However, it is well known that the OLS fixed-effect estimator overfits and induces a negative bias on the correlation of fixed effects. Hence, asymptotic small-sample bias correction techniques have been developed. Andrews et al. (2008) were the first to address this issue. In recent work, Kline et al. (2020) make the bias correction technique more practical by introducing a Jackknife technique and Azkarate-Askasua and Zerecero (2019) develop a bootstrap approach. In all cases, the bias correction is quite small. The residual explains less than 20% of the log-wage variance and the worker effect is nearly always explaining more than 50%. Finite-sample corrections presumably reduce the amount of overfitting but cannot eliminate it.[25]

The only estimation of an AKM model yielding a large residual component is the one in Bagger et al. (2013) on the same Danish data as ours. We thus performed Monte Carlo simulations of our estimated discrete mixture model. The artificial OLS estimates are in line with the variance decomposition estimated by Bagger et al. (2013). The residual contribution was reduced by about 10 percentage points, and the worker effect increased by 10–15 points. The increased firm effect and the reduced contribution of the covariance vary in differing proportions depending on the period. So, whether by estimating an AKM

[25]The discrete mixture model is similar to a LASSO penalization method designed to solve the ill-posedness of the latent-variable estimation problem. It can be viewed as a restriction on the number of different values that fixed effects can take. In our framework, it is a lot more than that because it is also a model of employment mobility that allows for more flexible interactions between heterogeneity and time-varying observables.

Table 8: Log-wage variance decompositions (in the literature)

		AKM													BLM
		FR	US1	AU	US2	IT	BR	DE		US3		DK	VE	FR2	SW
		76-87	84-93	90-97	90-99	81-97	95-05	85-91	02-09	80-86	07-13	87-13	99-01	02-14	02-04
Worker effect	Va	77	82	66	64	44	60	61	51	48	53	43	61	56	60
Firm effect	Vb	30	19	37	15	13	27	19	21	16	12	12	13	15	3
Residual	Eσ^2	16	9	5	9	15	7	8	5	20	15	39	10	16	25
Sorting effect	2Cov(a, b)	-27	-2	-22	1	2	3	2	16	2	7	3	16	-6	12
Match effect	V$\widetilde{\mu}$				5			3	2						
Observed heterogeneity	$V\overline{\mu}$	7	52	3	4	8	3	11	3	8	7	2		7	
	2Cov(a, x)	-3	-69	9	1	15		-3	1			1		10	
	2Cov(b, x)	1	9	2	1	3		2	2			1		1	
Sorting	Corr(a, b)	-.28	-.025	-.23	.010	.044	.040	.034	.25	.029	.14	.074	.283	-.11	.49

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. (2018). 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 VE is our own calculations based on Table II of Kline et al. (2018). They apply their "leave-one-out" estimator to Italian data from the Veneto region. Column FR2 is from Table 2 of Azkarate-Askasua and Zerecero (2019) and uses French data. Column BLM refers to Bonhomme et al. (2017)'s estimation for Sweden, 2002-04 (Table 2).

model on the actual or simulated data, it seems that the Danish data are special in that they display a greater idiosyncratic variance and a smaller contribution of worker types.

Finally, unless still biased downward, our estimation of the fixed effect correlation (between .2 and .3) is not a large one. The interpretation that we propose in Section 6 is that the correlation between worker and firm effects is a measure of sorting only if workers and firms can be classified along one single dimension, which would then, in turn, correlate with wages. If firms differ along several dimensions, with loading factors that also differ across countries, then fixed effect correlations may just become meaningless.

D.1 Conditional variance decomposition

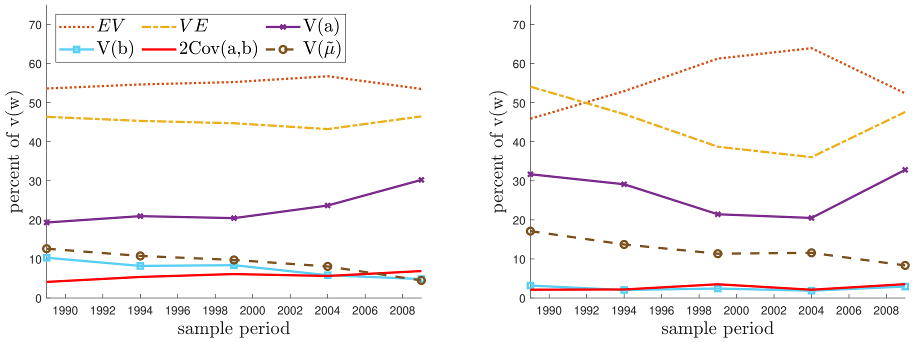
To understand the relative importance of interactions between tenure and experience, on the one hand, and worker and firm types, on the other, we repeat the variance decomposition analysis separately for each tenure-experience group. Figure 9 shows the results. At low tenure (left panel), there is higher residual variance, a bigger contribution of the firm effect, and a much smaller contribution of worker heterogeneity. There are also more interactions between worker and firm types—as reflected in a higher contribution of the sorting effect (the covariance of fixed effects). Overall, the relative importance of each factor, particularly worker effects, differs considerably by tenure status. This conditional variance decomposition shows how artificial certain results were; for instance, the huge contribution of tenure and experience in the last period. It seems very important to interact tenure and worker types.[26]

[26]This result is reminiscent of Guvenen (2009)'s comparison of HIP ("heterogeneous income profiles," ie heterogeneous returns to experience) and RIP ("restricted income profiles," ie income processes with a strong random walk component) models.

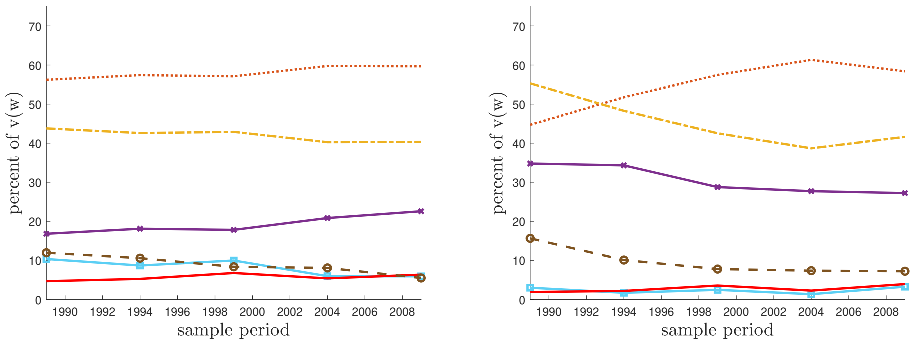
Figure 9: Conditional variance decomposition

Short tenure   Long tenure

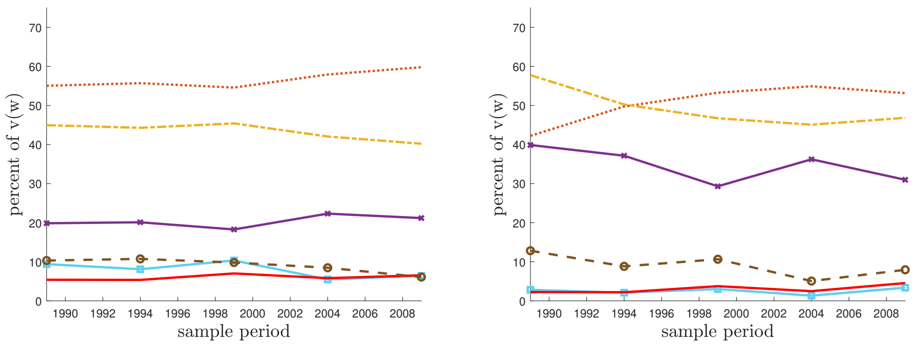
(a) Less than 5 years of experience



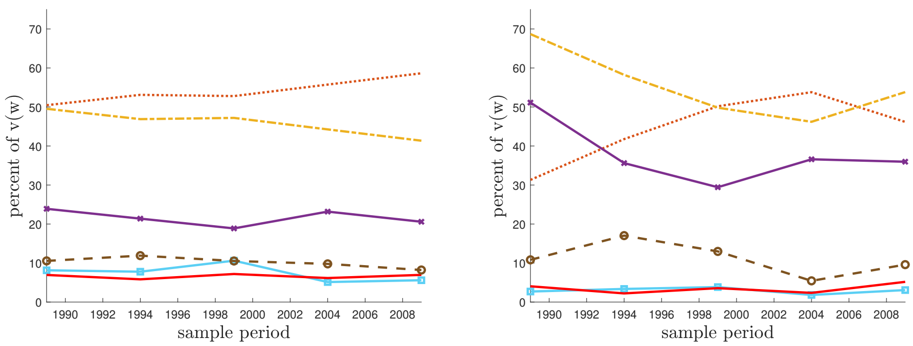
(b) 5-10 years of experience



(c) 10-15 years of experience



(d) At least 15 years of experience



E Additional results (online appendix)

Table 9: Firm characteristics by type in period 2

ℓ	no info	public	private	avg spell all ℓ-firms	no. firm	avg size/yr	avg inflow/yr	avg outflow/yr	avg age in yr
1	0.02	0.33	0.65	7960	6226	1.28	0.72	0.70	4.94
2	0.02	0.40	0.57	23369	4781	4.89	0.58	0.53	8.47
3	0.03	0.25	0.72	12526	9233	1.36	0.62	0.60	5.08
4	0.02	0.31	0.67	66569	7073	9.41	0.52	0.46	9.45
5	0.04	0.06	0.90	20176	14742	1.37	0.77	0.76	3.94
6	0.02	0.10	0.88	31959	20378	1.57	0.49	0.48	6.78
7	0.00	0.04	0.95	47013	22204	2.12	0.49	0.47	9.97
8	0.23	0.50	0.27	282366	150	1882.44	0.36	0.03	12.91
9	0.02	0.17	0.81	434100	5868	73.98	0.39	0.25	11.29
10	0.00	0.11	0.89	102101	19693	5.18	0.35	0.31	12.04
11	0.02	0.04	0.95	94658	14175	6.68	0.40	0.37	6.89
12	0.03	0.06	0.92	34998	31460	1.11	0.47	0.46	7.00
13	0.03	0.04	0.93	69139	22599	3.06	0.53	0.52	4.47
14	0.04	0.03	0.92	102721	5182	19.82	0.50	0.41	5.28

Table 10: Firm characteristics by type in period 3

ℓ	no info	public	private	avg spell all ℓ-firms	no. firms	avg size/yr	avg inflow/yr	avg outflow/yr	avg age in yr
1	0.05	0.33	0.62	4662	3778	1.23	0.76	0.70	6.36
2	0.05	0.55	0.41	4464	666	6.70	0.71	0.56	11.88
3	0.06	0.28	0.66	5221	5168	1.01	0.87	0.83	5.74
4	0.05	0.12	0.83	11679	4685	2.49	0.72	0.64	5.98
5	0.08	0.09	0.83	22360	300	74.53	0.74	0.33	7.00
6	0.02	0.46	0.51	12979	7392	1.76	0.64	0.55	12.24
7	0.07	0.10	0.83	11761	9475	1.24	0.85	0.82	5.25
8	0.02	0.66	0.32	26760	2994	8.94	0.55	0.43	15.31
9	0.05	0.06	0.89	49036	6651	7.37	0.65	0.55	7.37
10	0.06	0.06	0.88	23153	12956	1.79	0.56	0.54	6.52
11	0.22	0.56	0.23	285184	200	1425.92	0.39	0.02	16.48
12	0.00	0.12	0.88	41711	14645	2.85	0.35	0.31	15.07
13	0.11	0.07	0.82	71019	15276	4.65	0.42	0.38	7.99
14	0.01	0.22	0.78	237283	6292	37.71	0.41	0.25	14.94
15	0.02	0.06	0.92	25993	14381	1.81	0.53	0.49	11.22
16	0.10	0.04	0.86	36182	22943	1.58	0.47	0.46	6.97
17	0.04	0.02	0.95	29515	8721	3.38	0.72	0.62	7.02
18	0.07	0.02	0.91	120541	11822	10.20	0.40	0.32	9.20
19	0.03	0.05	0.92	94202	11143	8.45	0.52	0.44	9.77
20	0.06	0.03	0.90	15804	14494	1.09	0.79	0.77	5.71
21	0.10	0.03	0.87	220533	1949	113.15	0.42	0.19	10.86
22	0.15	0.03	0.82	41427	16187	2.56	0.54	0.51	5.49

Table 11: Firm characteristics by type in period 4

ℓ	no info	public	private	avg spell all ℓ-firms	no. firms	avg size/yr	avg inflow/yr	avg outflow/yr	avg age in yr
1	0.16	0.38	0.46	3895	3075	1.27	0.79	0.71	8.64
2	0.18	0.23	0.60	6941	5571	1.25	0.84	0.80	6.85
3	0.16	0.49	0.35	7423	1157	6.42	0.73	0.60	14.34
4	0.22	0.21	0.57	10819	6572	1.65	0.60	0.56	8.11
5	0.32	0.03	0.65	47178	7471	6.31	0.66	0.59	7.46
6	0.19	0.10	0.71	13240	9426	1.40	0.85	0.83	5.39
7	0.01	0.69	0.30	18706	4593	4.07	0.52	0.45	19.52
8	0.35	0.08	0.58	33863	12731	2.66	0.55	0.53	6.58
9	0.06	0.12	0.82	8408	10803	0.78	0.67	0.64	11.32
10	0.07	0.07	0.86	23614	11933	1.98	0.51	0.48	13.92
11	0.37	0.09	0.54	62922	18688	3.37	0.37	0.36	9.65
12	0.82	0.06	0.12	433508	94	4611.78	0.45	0.01	10.24
13	0.00	0.43	0.57	174620	1353	129.06	0.39	0.17	20.40
14	0.00	0.13	0.87	47447	11554	4.11	0.37	0.33	17.75
15	0.30	0.06	0.63	106902	10459	10.22	0.55	0.48	10.06
16	0.34	0.04	0.62	49422	24507	2.02	0.45	0.44	6.67
17	0.23	0.06	0.72	375017	5034	74.50	0.44	0.29	13.67
18	0.17	0.01	0.81	32934	7711	4.27	0.70	0.63	7.48
19	0.38	0.02	0.60	132247	12475	10.60	0.40	0.36	9.64
20	0.18	0.04	0.78	21437	15244	1.41	0.75	0.74	4.79
21	0.38	0.01	0.61	38556	11903	3.24	0.40	0.38	8.62
22	0.33	0.03	0.64	20837	12042	1.73	0.66	0.65	4.00

Table 12: Firm characteristics by type in period 5

ℓ	no. firms	no. workers	avg size/yr	legal status private	public	mixed	avg inflow/yr	avg outflow/yr	avg age
1	5442	5655	1.04	0.60	0.07	0.33	0.79	0.77	7.71
2	2425	32555	13.42	0.62	0.01	0.37	0.69	0.58	7.52
3	7091	20681	2.92	0.70	0.01	0.29	0.66	0.61	7.14
4	4866	8058	1.66	0.44	0.36	0.20	0.60	0.53	17.23
5	9067	12432	1.37	0.66	0.02	0.32	0.86	0.84	5.95
6	2554	18154	7.11	0.36	0.62	0.02	0.52	0.44	22.79
7	10150	70578	6.95	0.58	0.01	0.41	0.54	0.48	9.59
8	13781	25247	1.83	0.71	0.02	0.27	0.52	0.51	7.42
9	126	347968	2761.65	0.12	0.07	0.81	0.37	0.02	14.80
10	11313	18457	1.63	0.81	0.13	0.06	0.47	0.45	16.19
11	20705	54578	2.64	0.56	0.01	0.43	0.33	0.32	10.95
12	2200	213090	96.86	0.48	0.03	0.49	0.45	0.26	12.35
13	6669	16025	2.40	0.74	0.02	0.24	0.71	0.67	8.13
14	8519	32806	3.85	0.79	0.20	0.01	0.35	0.31	21.73
15	22301	43666	1.96	0.72	0.00	0.27	0.46	0.45	7.00
16	6781	18851	2.78	0.67	0.00	0.32	0.71	0.64	9.07
17	10267	101446	9.88	0.60	0.01	0.38	0.50	0.44	11.21
18	14599	15193	1.04	0.65	0.01	0.34	0.79	0.78	5.82
19	3408	113882	33.42	0.80	0.20	0.00	0.36	0.23	23.65
20	11098	83863	7.56	0.55	0.01	0.44	0.36	0.32	10.78
21	15429	20233	1.31	0.63	0.00	0.37	0.52	0.52	7.31
22	5368	24685	4.60	0.71	0.00	0.28	0.45	0.41	13.63

Table 13: Period 5: 2009-2013