A Distributional Framework for Matched Employer Employee Data *

Stéphane Bonhomme
University of Chicago

Thibaut Lamadon
University of Chicago

Elena Manresa
MIT Sloan

First version: November, 2014
This version: March 17, 2015

PRELIMINARY AND INCOMPLETE

Abstract

We develop a discrete heterogeneity framework for matched employer employee data. The framework allows for unrestricted interactions between worker and firm unobserved characteristics in the wage function, as well as unrestricted sorting based on these unobservables. Pooling cross-sectional observations together with information from the joint distribution of wages of job movers, we establish a series of nonparametric identification results in short panels. We evaluate our method on data simulated from a theoretical model under both positive and negative sorting. We apply our method to Swedish matched employer employee panel data and report estimated wage functions and sorting patterns.

JEL codes: J31, J62, C23.

Keywords: matched employer employee data, sorting, job mobility, models of heterogeneity.

*We thank Rafael Lopes de Melo, Costas Meghir, Luigi Pistaferri, Jean-Marc Robin and Robert Shimer for useful comments. We thank the IFAU for access to, and help with, the Swedish administrative data. Contact: sbonhomme@uchicago.edu; lamadon@uchicago.edu; emanresa@mit.edu .

1 Introduction

Identifying the contributions of worker and firm heterogeneity to wage dispersion is an important step towards answering a number of economic questions, such as the nature of sorting patterns between heterogeneous workers and firms, the optimal allocation of workers to firms, or the sources of wage inequality.

To this end, for more than twenty years researchers have relied on matched employer employee data, where workers are followed over time and across firms, and the data contain both worker and firm identifiers. This provides the opportunity to allow for unobserved worker and firm heterogeneity, in addition to heterogeneity coming from covariates such as education, age, or firm size, for example.

In an influential article, Abowd, Kramarz, and Margolis (1999) (AKM hereafter) developed a method to recover worker and firm effects on wages using matched data. In a regression framework, they proposed to estimate worker and firm fixed-effects, thus allowing to quantify their respective contributions to wage dispersion, as well as the correlation between worker and firm unobservables. Applications of the method to wage data include Abowd, Kramarz, Pérez-Duarte, and Schmutte (2009), Gruetter and Lalive (2009), Mendes, van den Berg, and Lindeboom (2010), Woodcock (2008), and Card, Heining, and Kline (2013), among others. The AKM method has been used in a variety of other fields, for example to link banks to firms, or teachers to schools or students (Kramarz, Machin, and Ouazad, 2008; Jackson, 2013).

Since the publication of AKM, the literature has emphasized several challenges to the use of two-way fixed-effects regressions. One feature of the method is that the model of wages, in logarithms, is additive in worker and firm effects. This rules out the possibility that firm-specific returns might be heterogeneous across workers, and imposes tight restrictions on patterns of complementarity between workers and firms. In addition to being possibly rejected empirically, additivity may also be inconsistent with theoretical models of sorting (e.g. Shimer and Smith, 2000; Eeckhout and Kircher, 2011). Another challenge with the AKM method concerns estimation, as fixed-effects estimates may be poorly estimated even in panels that follow workers and firms for long periods of time, possibly leading to biased estimates of sorting (Andrews, Gill, Schank, and Upward, 2008, 2012).

In this paper we take a different approach and propose a distributional framework for matched data. The approach allows for unrestricted interaction effects between firm and worker unobservables, and it provides estimates of conditional wage distributions given worker and firm heterogeneity, as well as estimates of sorting of workers across firms. Moreover, the method can

deal with short panels with as few as two periods. This is important, since it opens the way to document how wage returns and worker sorting vary over the business cycle, for example. In addition, although we focus on workers and firms in this paper, our framework could be useful in other applications, such as teacher-student sorting, where long panels may not be available.

In order to make the matched problem tractable, we rely on an asymmetric modeling of firm and worker heterogeneity. We model both types of unobservables as discrete. Discreteness is convenient for implementation, although we also discuss extensions to models with continuously distributed unobservables. On the firm side, unobservables are specified as belonging to a finite number of latent classes. This discrete fixed-effects approach has recently been proposed in single-agent panel data analysis (Hahn and Moon, 2010; Lin and Ng, 2012; Bonhomme and Manresa, 2012). With sufficiently many workers per firm, firm membership to the different classes will be accurately estimated. In contrast, the number of observations for a given worker is typically small. For this reason, we use a random-effects approach to model the distribution of latent worker types within each firm class.

In the model there are two reasons why firms that belong to different classes differ from each other. First, firm heterogeneity shows up in worker composition, which depends on the firm class. Estimating distributions of worker heterogeneity within each class will thus reveal the extent of sorting in the economy. Second, firms in different classes have different conditional wage distributions given worker types, for example as a result of differences in productivity or technology. Recovering conditional wage distributions will allow us to document complementarity between firm and worker heterogeneity. The first feature links our model to "mixed membership models" that have recently become popular in machine learning and statistics (Blei, Ng, and Jordan, 2003; Airoldi, Blei, Fienberg, and Xing, 2008). However, as far as we know the combination of the two features is novel to our setup.

In order to nonparametrically identify the model, we rely on workers who move between firms. This variation is also central to the AKM methodology. In the main analysis we make two assumptions that are analogous to assumptions made by AKM: we assume that job mobility does not depend on past wages given the type of the worker and the latent classes of the firms before and after the move (we refer to this assumption as "exogenous mobility"), and that wages in a firm after a job move do not depend on the previous wage, given the worker type and the firm unobservable class ("serial independence"). Under these two assumptions, we show that a bivariate distribution of wages before and after a move, together with a cross-section of wages, are sufficient for identification. In addition, we also show that, when an additional wage observation is available before and after the job move, both endogenous mobility and serial

correlation can be allowed for.

The model is a conditional finite mixture, where the conditioning variables are the latent firm classes. These models are generally nonparametrically under-identified based on two periods of observation only (Hall and Zhou, 2003; Henry, Kitamura, and Salanié, 2014). We exploit job movements across different firm classes and show that, in our framework, two periods (that is, a single job move per worker) are sufficient for recovering type- and class-specific wage distributions as well as the type composition of job movers within each pair of firm classes, before and after the move. Intuitively, the wages of movers to a given firm class who come from different classes provide variation in terms of worker type composition. This variation allows us to disentangle the impact of differences in worker composition from that of differences in conditional wage distributions.

Recovering conditional wage distributions requires identifying class membership for each firm. To do so, we use the fact that firm-specific wage distributions are in fact class-specific. This allows us to directly identify class membership from the joint distribution of wages. Lastly, having shown that the latent firm classes and the wage distributions are identified, we show how to recover the proportions of latent worker types within each firm class in the cross-section.

The identification arguments suggest a three-step approach to estimation: clustering firms into classes, and estimating finite mixture models to recover conditional wage distributions from the subsample of job movers, and worker type proportions from the cross-section, in turn. Clustering firms can be done using the k-means algorithm (Steinley, 2006) applied to firm-specific empirical cumulative distribution functions. The finite mixture models can then be estimated under parametric assumptions or using non-parametric methods.

An influential body of work, built on Becker (1974), has developed and estimated fully specified theoretical models of sorting on the labor market (De Melo, 2009; Lise, Meghir, and Robin, 2008; Bagger, Fontaine, Postel-Vinay, and Robin, 2011; Hagedorn, Law, and Manovskii, 2014; Lamadon, Lise, Meghir, and Robin, 2013; Bagger and Lentz, 2014). The aim of this work is to address fundamental questions such as the efficiency of the allocation of worker to jobs and how it is affected by policy. Without further assumptions our framework does not directly address these questions. Nevertheless, as we neither restrict the complementarity between worker and firm heterogeneity, nor the sorting patterns, results based on our framework can inform this structural literature. To illustrate the connection between a structural approach and ours, we provide an explicit mapping between our model and an extension of the model of Shimer and Smith (2000) with on-the-job search, and we document the performance of our estimation method on data generated according to the theoretical model.

We take our approach to Swedish matched employer employee panel data for 1997-2006. While the preliminary evidence that we report in this version of the paper suggests some departure from additivity between firm and worker heterogeneity, we find that additive models approximate the conditional mean of wages relatively well. At the same time, we find substantial sorting of workers across firms, mostly in terms of unobservables, and we show that the AKM methodology does not recover the amount of sorting in our data.

The outline of the paper is as follows. In Section 2 we present the framework of analysis. In Sections 3 and 4 we study identification and estimation, respectively. In Sections 5 and 6 we describe the data and show preliminary empirical results. In Section 7 we analyze several extensions of the main model, including how to allow for endogenous mobility, serial correlation, and continuously distributed heterogeneity. Lastly, we conclude in Section 8.

2 Framework of analysis

2.1 Economic environment

We consider an economy composed of workers indexed by i with discrete types $\omega(i) \in \{1, ..., K\}$, and firms indexed by j with discrete classes $f(j) \in \{1, ..., L\}$. Worker types and firm classes may reflect differences in productivity or other unobservable traits. We observe this economy over T periods. In this section we take $T = 2$. We outline an extension of the model to $T > 2$ periods in Section 7.

Period-1 wages. Workers are either employed in a firm or unemployed. We denote by j_{it} the identifier of the firm where i, if employed, works at time t. In period 1, which corresponds to the start of the observation period, employed workers earn a log wage Y_{i1}, which may depend on worker type and firm class. We denote by $F_{k\ell}$ the corresponding log wage cumulative distribution function (cdf):

$$F_{k\ell}(y) := \Pr\left[Y_{i1} \leq y \,|\, \omega(i) = k, f(j_{i1}) = \ell\right], \tag{1}$$

for all $k \in \{1, ..., K\}$ and $\ell \in \{1, ..., L\}$. In (1) we abstract from the effect of observed covariates. We return to this issue in Section 4.

Initial allocation. In period 1, worker types are drawn within firm from a multinomial distribution indexed by the firm class. The proportion of type-k workers in firm j is given by

$\pi_k(f(j))$, where $\pi_k(\ell)$ is defined as:

$$\pi_k(\ell) := \Pr\left[\omega(i) = k \,|\, f(j_{i1}) = \ell\right]. \tag{2}$$

Similarly, the proportion of type k workers in unemployment is given by $\pi_k(0)$.

In period 1 firms differ in terms of worker composition. This allows for sorting without restricting its pattern. For example, both positive and negative assortative matching are possible. The model thus shares similarities with mixed membership models (Blei, Ng, and Jordan, 2003; Airoldi, Blei, Fienberg, and Xing, 2008). One difference with these models is that here firms do not only differ in terms of the type of workers they employ, but also in terms of the wages they offer to workers of a given type. To separately identify both sources of heterogeneity we rely on panel data.

Job mobility. Workers employed in period 1 can either stay in their job, move to another job, or become unemployed in the next period. In the main analysis we make two assumptions regarding job mobility. The first assumption is that job mobility is only driven by worker types and firm classes, not directly by wages. We will refer to it as a condition of "exogenous mobility", consistently with the literature following Abowd, Kramarz, and Margolis (1999). Formally, let $m_i := \mathbf{1}\{j_{i1} \neq j_{i2}\}$ denote the indicator that worker i changes jobs between period 1 and 2.

Assumption 1. *(exogenous mobility)*

$$Y_{i1} \text{ is independent of } (m_i, f(j_{i2})) \text{ given } (\omega(i), f(j_{i1})).$$

The second assumption, "serial independence", is that wages drawn from different firms are conditionally independent over time, given worker type and firm class.

Assumption 2. *(serial independence)*

$$Y_{i2} \text{ is independent of } (Y_{i1}, f(j_{i1})) \text{ given } (\omega(i), f(j_{i2}), m_i = 1).$$

We will comment on Assumptions 1 and 2 in the next subsection. Moreover, in Section 7 we will propose a four-period extension of the model where we will relax both assumptions.

Period-2 wages of job movers. We also assume that wages of job movers in period 2, conditional on the firm class and the worker type, are drawn from the same conditional distribution as period-1 wages.

Assumption 3. *(stationarity)*

$$\textit{for all } k, \ell' : \qquad \Pr\left[Y_{i2} \leq y' \,|\, \omega(i) = k, f(j_{i2}) = \ell', m_i = 1\right] = F_{k\ell'}(y').$$

In our two-period setup it is possible to relax Assumption 3. For example, it may be important to allow for period-specific log wage distributions $F^t_{k\ell}$, in order to capture macroeconomic shocks or life-cycle effects. On might also wish to allow job movers to earn different wages than the rest of the employed population, beyond type differences. Another extension which our approach covers is when firm classes $f^t(j)$ vary over time, for example as a result of productivity shocks at the firm level. We analyze both extensions in Section 7. In the empirical illustration we will work under Assumption 3, while netting log wages of period dummies.

Under Assumptions 1, 2 and 3 we have, for all k, ℓ, ℓ':

$$\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, \omega(i) = k, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] = F_{k\ell}(y) F_{k\ell'}(y'). \tag{3}$$

Type composition of job movers. Similarly as for the initial worker allocation, we do not restrict the type composition of job movers. That is, for moves taking place between period 1 and 2, we define, for all k, ℓ, ℓ':

$$p_k\left(\ell, \ell'\right) := \Pr\left[\omega(i) = k \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right], \tag{4}$$

where $p_k\left(\ell, \ell'\right)$ are unrestricted proportions.

This means that, while strict exogeneity in Assumption 1 holds, job mobility may depend on worker type, and on both firm classes before and after the move, in an unrestricted way.

Model's main restrictions. To recover $F_{k\ell}$, $p_k(\ell, \ell')$, and $\pi_k(\ell)$, it is sufficient to focus on the first two periods, and exploit the following two equations:

$$\Pr\left[Y_{i1} \leq y\,,\, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] = \sum_{k=1}^{K} p_k\left(\ell, \ell'\right) F_{k\ell}(y) F_{k\ell'}(y'), \tag{5}$$

and:

$$\Pr\left[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell\right] = \sum_{k=1}^{K} \pi_k\left(\ell\right) F_{k\ell}(y). \tag{6}$$

Equation (5) represents the bivariate wage distribution of job movers conditional on both firm classes. It will allow us to recover $F_{k\ell}$ and $p_k(\ell, \ell')$. Equation (6) represents the univariate cross-sectional wage distribution in period 1. It will allow us to recover $\pi_k(\ell)$. In practice, one

could supplement (6) with the cross-sectional wage distribution in period 2, although we do not impose that worker composition $\pi_k(\ell)$ remains constant between the two periods.

Our main identification result, which we establish in Section 3, is that, under suitable rank conditions, $F_{k\ell}$ and $p_k(\ell, \ell')$ are identified from (5), and $\pi_k(\ell)$ are then identified from (6). Moreover, the partition of firms into classes is also identified from (5) and/or (6).

2.2 Discussion

Discussion of Assumptions 1 and 2. To illustrate the requirements contained in Assumptions 1 and 2, consider a linear model similar to the one in Abowd, Kramarz, and Margolis (1999):

$$Y_{it} = \alpha(i) + \delta(j_{it}) + \varepsilon_{it}, \tag{7}$$

where we have abstracted from observed covariates. Denote $j_i = (j_{i1}, ..., j_{iT})$. The condition of exogenous mobility in AKM is $\mathbb{E}(\varepsilon_{it} \,|\, j_i) = 0$. Analogously, in Assumption 1 we require that the wage in period 1 does not affect job mobility directly, conditional on worker types and firm classes. The condition of serial uncorrelatedness in AKM is $\mathbb{E}(\varepsilon_{it}\varepsilon_{is} \,|\, j_i) = 0$ for $t \neq s$. In Assumption 2 we similarly require that wages in period 2 do not depend on wages in period 1, given worker types and firm classes. Note that, while like AKM our framework imposes serial independence across jobs, it leaves serial correlation within jobs unrestricted.

Although it is possible to allow for serial correlation in AKM, at least in a long panel setting, strict exogeneity has proven more challenging to relax. Given this, the extensions in Section 7 are of particular interest. Nevertheless, working under Assumptions 1 and 2 has the advantage of implying a simple structure on bivariate wage distributions conditional on firm classes before and after a move, and of allowing for identification based on two periods only.

Quantities of interest. In this paper our interest mainly centers on two quantities: the type- and class-specific wage distributions, $F_{k\ell}$, and the composition of each firm class in terms of worker types, $\pi_k(\ell)$. The latter quantity is informative about sorting, while the former captures how wages depend on worker and firm heterogeneity.

With these two objects at hand one can generate counterfactual allocations. For example, one can describe the outcome of random matching, obtained by imposing that $\pi_k(\ell)$ is independent of ℓ. Alternatively, taking the perspective of the worker, one could assign each $\omega(i)$ to the firm j such that the mean of $F_{\omega(i),f(j)}$ is maximum, or one could assign workers to firms in a way that maximizes the sum of wages in the economy. Other possible exercises include variance decompositions, and more generally distributional wage decompositions. Also, one can compare

the predictive power of linear combinations of worker and firm types in wage regressions, with fully saturated specifications that allow for unrestricted interactions.

In addition, our setup allows to recover the composition of each pair of firm classes in terms of worker types within the set of job movers, $p_k(\ell, \ell')$. These proportions are useful to document dynamic sorting patterns as, by Bayes' rule:

$$\Pr\left[f(j_{i2}) = \ell' \mid f(j_{i1}) = \ell, \omega(i) = k, m_i = 1\right] = \frac{q_{\ell\ell'}\, p_k(\ell, \ell')}{\sum_{\widetilde{\ell}=1}^{L} q_{\ell\widetilde{\ell}}\, p_k(\ell, \widetilde{\ell})}, \tag{8}$$

where $q_{\ell\ell'} = \Pr\left[f(j_{i2}) = \ell' \mid f(j_{i1}) = \ell, m_i = 1\right]$. These measures, which we will document in our empirical illustration below, complement the notion of cross-sectional sorting captured by $\pi_k(\ell)$.

Other transitions. While two periods and equations (5) and (6) are sufficient to identify the quantities we focus on in this paper, it is possible to extend the framework to a complete model of labor market transitions and wages over T periods. For this, one needs to model the wage distributions and dynamics for job stayers, and transitions in and out of unemployment. See Section 7 for details.

Links to theoretical models. Our setup allows for general forms of complementarity between worker types and firm classes. This feature is important as, since Becker (1974), the theoretical and structural literature has emphasized the close link between complementarity or substitution patterns and the amount of sorting. See, for example, Shimer and Smith (2000), Eeckhout and Kircher (2011), and Hagedorn, Law, and Manovskii (2014).

As our model neither restricts complementarity nor sorting patterns, it provides a general framework that can be useful to inform structural models. In Appendix B we provide an explicit mapping between an extension of the search-matching model of Shimer and Smith (2000) with on-the-job search and our framework. We also show through simulations that, when the data are generated according to the theoretical model, our method recovers the underlying wage structure quite accurately. The quality of approximation is only moderately affected when the data generating process has a large number of heterogeneous types but, as in the empirical application, we use a small number of types for estimation.

In models which build on Shimer and Smith (2000), both worker and firm productivity types are ordered in the population. Our method, which is based on wage data, does not directly allow us to recover the ordering. In practice, one may learn about the latent types and classes by documenting how they affect wages and correlate with observed covariates. Moreover, log

wage variance decomposition exercises, such as the ones we report in the empirical analysis below, do not depend on recovering a productivity ordering. In fact, our approach does not require worker or firm heterogeneity to be unidimensional.

In a number of theoretical models of sorting, strict exogeneity in Assumption 1 is not satisfied. This will be the case if job mobility depends on previous wages, in addition to the type of the worker and the firm classes. To relax strict exogeneity, one needs to model how mobility to a class-ℓ' firm depends on the previous firm class, worker type, and the previous wage. In Section 7 we show how the availability of an additional wage observation before and after a job move allows to identify this more general model.

Lastly, while our model and its extension that allows for endogenous mobility nest a number of theoretical models of sorting, they do not specify the entire economic structure, such as production, outside options, or surplus. As a result we will not be able to perform an exhaustive set of counterfactual exercises. Nevertheless, as our empirical application will illustrate, our framework is useful to analyze the dependence of the wage structure on worker and firm heterogeneity, and to document the effect of sorting on wage dispersion.

3 Identification

In this section we start by showing how, for a given partition of firms into classes and under suitable conditions, $F_{k\ell}$, $\pi_k(\ell)$, and $p_k(\ell, \ell')$ are all identified from (5) and (6). Then we show how class membership $f(j)$, for each firm j, is also identified.

3.1 Intuition in a simple model

To provide an intuition for why the distributions $F_{k\ell}$ are identified, we first consider the following simple model:

$$Y_{it} = a(f(j_{it})) + b(f(j_{it}))\alpha_i + \sigma(f(j_{it}))\varepsilon_{it}, \tag{9}$$

where ε_{it} is independent of α_i and $f(j_{it})$, distributed as a standard normal, and independent over time. The partition of firms into classes is assumed known. α_i has unrestricted mean and variance given firm class. When α_i is discrete with K points of support, model (9) is a special case of the framework described in Section 2. Note that here α_i could be continuously distributed, for example Gaussian. We outline an extension of our framework to continuous worker types in Section 7.

Consider job movers between classes ℓ and $\ell' \neq \ell$, between period 1 and 2. We have:

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

where, conditional on the transition from ℓ to ℓ', α_i is normally distributed with mean and variance $\mathbb{E}_{\ell\ell'}(\alpha_i)$ and $\text{Var}_{\ell\ell'}(\alpha_i)$, respectively, and where (without loss of generality) we have normalized $a(\ell') = 0$ and $b(\ell') = 1$.

This model is formally equivalent to a measurement error model where α_i is the error-free regressor and Y_{i2} is the error-ridden regressor. It is well-known that $a(\ell)$, $b(\ell)$ and $\sigma(\ell)$ are not identified using mean and covariance restrictions only. As an example, identification fails when α_i is Gaussian (Reiersøl, 1950).

Now consider also job movers from class ℓ' to class ℓ. We have:

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

In this case one can show that $a(\ell)$, $b(\ell)$ and $\sigma(\ell)$ are identified, provided:

$$\mathbb{E}_{\ell\ell'}(\alpha_i) \neq \mathbb{E}_{\ell'\ell}(\alpha_i). \tag{10}$$

For example, we have:

$$b(\ell) = \frac{\mathbb{E}_{\ell'\ell}(Y_{i2}) - \mathbb{E}_{\ell\ell'}(Y_{i1})}{\mathbb{E}_{\ell'\ell}(Y_{i1}) - \mathbb{E}_{\ell\ell'}(Y_{i2})}.$$

Moreover, the means and variances of α_i in the two firm classes and the two periods are also identified.

An intuition for these results, which is also at the core of our main identification result in the next subsection, is as follows. For a given worker type, the wage gains or losses associated with job mobility are due to differences in the firm wage schedules, represented by the a, b and σ parameters. Averaging over types, observed wage differences also reflect the effect of worker composition. If type composition is different between $\ell \to \ell'$ and $\ell' \to \ell$ job moves, for example, this variation can be used to identify the effect of worker heterogeneity on firm wages. Equation (10) requires type composition to differ between the two transitions. Note that, if $b(\ell) \neq -1$ this condition is equivalent to:

$$\mathbb{E}_{\ell'\ell}\left(Y_{i1} + Y_{i2}\right) \neq \mathbb{E}_{\ell\ell'}\left(Y_{i1} + Y_{i2}\right), \tag{11}$$

so it can be empirically tested. In the empirical sections we will show evidence supporting (11) in the Swedish data.

Note also that, under additivity (that is, if $b(\ell) = 1$):

$$\mathbb{E}_{\ell'\ell}\left(Y_{i2} - Y_{i1}\right) = \mathbb{E}_{\ell\ell'}\left(Y_{i1} - Y_{i2}\right), \tag{12}$$

which means that wage gains and losses associated with job changes between different firm classes are symmetric. Card, Heining, and Kline (2013) have exploited this idea to provide evidence suggesting that the additive AKM model is an appropriate specification on German data; see Figures Va and Vb in their paper. Here we use differences in worker type composition, reflected in (11), to relax additivity in estimation.

Finally, the fact that we cluster firms together into classes is useful, in the context of this example and also in the general discrete framework of Section 2, because it ensures that there is a large number of movers in both directions, $\ell \to \ell'$ and $\ell' \to \ell$. With very large firms and sufficient movements between those firms, the above identification argument could be conducted at the firm level, by taking ℓ and ℓ' to be firm identifiers. This observation could be useful in other contexts, such as movements of workers or firms between cities, where the number of movers (to or from a given city in that example) could be large.

3.2 Wage distributions and proportions

We are now in position to show that the distributions $F_{k\ell}$ are nonparametrically identified based on two periods of data with one job move, see equation (5). To show the result we start with a definition.

Definition 1. *A cycle of length R is a sequence of firm classes $(\ell_1, ..., \ell_R) \in \{1, ..., L\}^R$, with $\ell_{R+1} = \ell_1$, such that $p_k(\ell_r, \ell_{r+1}) \neq 0$ for all $r \in \{1, ..., R\}$ and $k \in \{1, ..., K\}$.*

Assumption 4. *(cycles)*
For any two firm classes $\ell \neq \ell'$ in $\{1, ..., L\}$, there exists a cycle of length $2R$ containing ℓ and ℓ' such that:
i) The scalars $a_1, ..., a_K$ are all distinct, where:

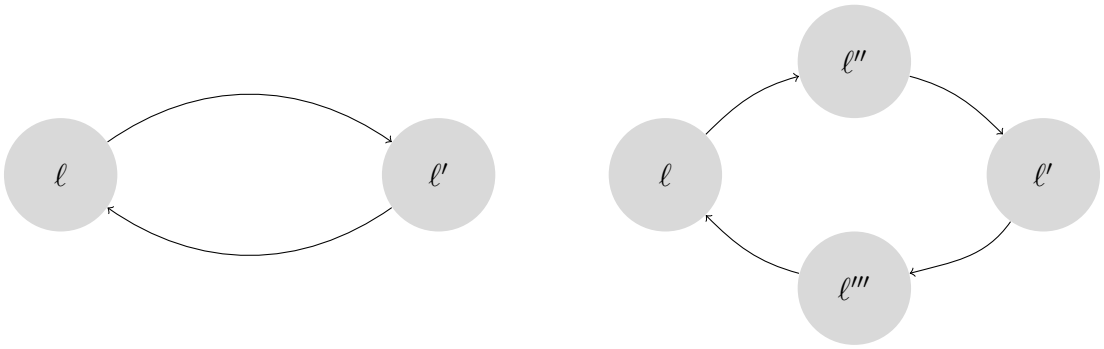
$$a_k := \frac{p_k(\ell_1, \ell_2)p_k(\ell_3, \ell_4)...p_k(\ell_{2R-1}, \ell_{2R})}{p_k(\ell_2, \ell_3)p_k(\ell_4, \ell_5)...p_k(\ell_{2R}, \ell_1)}.$$

ii) The matrices $A(\ell_1, \ell_2), ..., A(\ell_{2R}, \ell_1)$ have rank K, where:

$$A(\ell_r, \ell_{r+1}) := \left\{\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell_r, f(j_{i2}) = \ell_{r+1}, m_i = 1\right]\right\}_{(y,y')}$$

Assumption 4 requires that any two firm classes ℓ and ℓ' belong to a cycle of even length. An example is when there is a positive proportion of every worker type within the set of movers from ℓ to ℓ', and a positive proportion of every worker type within the set of movers from ℓ' to ℓ. Another example is when there are no job movers from ℓ to ℓ' or from ℓ' to ℓ, but some workers move from ℓ to ℓ'', others from ℓ'' to ℓ', some workers from ℓ' to ℓ''', and some workers move from ℓ''' to ℓ, as illustrated in Figure 1. Note that ℓ''' and ℓ' could coincide, for example, as the assumption allows for job movements with a class. Hence the requirement for the cycles to be of even length (which does not seem possible to extend to general odd-length cycles given our assumptions) does not appear to be a substantial limitation in practice. Existence of cycles is related to, but different from, that of connected groups in AKM (Abowd, Creecy, and Kramarz, 2002). As in AKM, in our setup identification will fail in the presence of completely segmented labor markets where firms are not connected between groups via job moves. One difference with AKM is that, in our nonlinear setup, we need every firm class to contain job movers of all types of workers.

Figure 1: Two cycles containing ℓ and ℓ'



The requirements on cycles can be relaxed, at the cost of loosing identification of some of the quantities of interest. To see this, let us assume that for all ℓ there exists an $\ell' \neq \ell$ such that $p_k(\ell', \ell) \neq 0$ for $k \in \mathcal{K}$, and $\pi_k(\ell) = 0$ (so $p_k(\ell', \ell) = p_k(\ell, \ell') = 0$) for $k \notin \mathcal{K}$. Moreover, let us suppose that the ratios $p_k(\ell, \ell')/p_k(\ell', \ell)$, for $k \in \mathcal{K}$, are all distinct. Then it can be shown using similar arguments as in the proof of Theorem 1 below that the cdfs $F_{k\ell}$ are identified for all $k \in \mathcal{K}$, although they are not identified for $k \notin \mathcal{K}$.

Part i) in Assumption 4 requires some asymmetry in worker type composition between different firm classes. This condition requires either non-random sorting or non-random mobility, as it fails when $p_k(\ell, \ell')$ does not depend on (ℓ, ℓ'). Another case where part i) fails is when cross-sectional sorting and sorting associated with job mobility exactly offset each other, so

$p_k(\ell, \ell')$ is symmetric in (ℓ, ℓ'). This exact offsetting happens in the model of Shimer and Smith (2000) in the absence of on-the-job search, as we discuss in Appendix B. In our framework the presence of asymmetric job movements between firm classes is crucial for identification. In the empirical analysis we will provide evidence of such asymmetry.

Lastly, part *ii*) is a rank condition. It will be satisfied if, in addition to part *i*), for all r the distributions F_{1,ℓ_r}, ..., F_{K,ℓ_r} are linearly independent. Note that this assumption is testable, and that it can provide guidance on the choice of K in practice.

The next result shows that, with only two periods and given the structure of the model, both the type-specific wage distributions and the type composition of job movers can be uniquely recovered. The intuition for the result is similar to the one we used in model (9).

Theorem 1. *Consider the joint distribution of wages of job movers, and let Assumptions 1, 2, 3 and 4 hold. Suppose that firm classes $f(j)$ are known, for all j. Then, up to common relabelling of the k index, $F_{k\ell}$ is identified for all (k, ℓ). Moreover, for all pairs (ℓ, ℓ') for which there are job moves from ℓ to ℓ' $p_k(\ell, \ell')$ is identified for all k up to the same relabelling.*

Proof. See Appendix A. ∎

Worker composition. Theorem 1 shows that the proportions of worker types, in the set of movers from a firm in class ℓ to another firm in class ℓ', are identified. We now show that the type proportions are also identified within each firm class in the cross-section in period 1.

Corollary 1. *Suppose that Assumptions 1, 2, 3 and 4 hold. Then $\pi_k(\ell)$ are identified up to a common relabelling of the k index.*

Proof. See Appendix A. ∎

3.3 Firm classes

The identification results in the previous subsection were conditional on knowing the partition of firms into classes. Here we show that this partition is also identified.

Identification based on job movers. Let j and j' be two different firms, which belong to classes $f(j) = \ell$ and $f(j') = \ell'$, respectively. The bivariate cdf of log wages conditional on moving from j to j' between period 1 and 2 is, by (5):

$$\begin{aligned} \Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] &= \sum_{k=1}^{K} p_k(\ell, \ell') F_{k\ell}(y) F_{k\ell'}(y') \\ &:= G_{\ell\ell'}(y, y'). \end{aligned} \tag{13}$$

This cdf only depends on the firm classes ℓ and ℓ', not directly on the firm identifiers j and j'.

Here we study identification of firm classes $f(j)$ assuming, for every firm j, that the number of movers from j to some other firm $j' \neq j$, or from some firm $j' \neq j$ to firm j, is infinite. In practice, this means that the number of job movers in or out of firm j needs to be large. Moreover, here we assume that the total number of firms J is finite. In Section 4 we will study asymptotic properties of our clustering estimator as both the number of firms and the number of workers tend to infinity, allowing the number of workers per firm to grow polynomially more slowly than the number of firms.

A sufficient condition for $f(j)$ to be identified for all j from (13) is as follows.

Assumption 5. *(separation) For all $\ell \neq \ell'$ in $\{1, ..., L\}$, there exists an $\ell'' \in \{1, ..., L\}$ such that either $G_{\ell\ell''} \neq G_{\ell'\ell''}$ or $G_{\ell''\ell} \neq G_{\ell''\ell'}$.*

Proposition 1. *Suppose that Assumptions 1, 2, 3 and 5 hold. Then the partition of firms $j \in \{1, ..., J\}$ into L classes $f(j) \in \{1, ..., L\}$ is identified.*

Proof. See Appendix A. ∎

Identification of the firm classes may fail if Assumption 5 does not hold for some $\ell \neq \ell'$. In that case, the next result shows that, under the conditions of Theorem 1, $F_{k\ell}$ and $F_{k\ell'}$ coincide for all k, which implies that grouping ℓ and ℓ' together does not result in a loss of information, as their conditional wage distributions given worker types are identical. Note that, due to symmetry between the two periods in the model, it is enough to assume that either of the two parts in Assumption 5 fails to hold in order to show the equality of the conditional wage distributions.

Corollary 2. *Let $\ell \neq \ell'$ such that $G_{\ell\ell''} = G_{\ell'\ell''}$ for all ℓ'', or $G_{\ell''\ell} = G_{\ell''\ell'}$ for all ℓ''. Suppose that the assumptions in Theorem 1 hold. Then $F_{k\ell} = F_{k\ell'}$ for all $k \in \{1, ..., K\}$.*

Proof. See Appendix A. ∎

Identification of classes based on a cross-section. The identification strategy in the previous paragraph relies on having a large number of job movers per firm. In practice this may be a strong requirement. An alternative approach is obtained by noticing that, for all j that belong to class $f(j) = \ell$, the cdf of log wages in firm j in period 1 is:

$$\begin{aligned} \Pr\left[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell\right] &= \sum_{k=1}^{K} \pi_k(\ell) F_{k\ell}(y), \\ &:= H_\ell\left(y\right), \end{aligned} \tag{14}$$

which only depends on the firm class $f(j) = \ell$. It is then easy to show that the firm partition into classes is identified from (14) if, for all $\ell \neq \ell'$, $H_\ell \neq H_{\ell'}$.

One difficulty with identifying firm classes from cross-sectional observations only is that it might be that two cross-sectional wage distributions coincide between two firms, one offering a higher wage schedule but having low-type workers, the other one offering a lower wage schedule but having high-type workers. This possibility has been emphasized in the theoretical literature (e.g. Eeckhout and Kircher, 2011). So it may be impossible to separate two different classes from the cross-section, even though their conditional wage distributions given worker types are different.

Corollary 2 shows that, provided that conditional wage distributions differ, information from the wages of job movers can allow to identify firm classes even when the cross-sectional wage information is insufficient. For this reason, when implementing the approach one may consider a method that clusters firm-specific distributions in the cross-section and adds information from the job movers, thus combining (13) and (14).

4 Estimation

Let N be the number of workers, and J the number of firms. The identification strategy suggests a three-step method to estimate the model's parameters. In the first step we estimate the firm classes $f(j)$ for all $j \in \{1, ..., J\}$ by clustering firm-specific distributions. In the second step, given firm classes, we estimate $F_{k\ell}$ and $p_k(\ell, \ell')$ from the finite mixture model in the sample of job movers. In the third step we recover the proportions $\pi_k(\ell)$ from the cross-section.

We now detail these three steps.

4.1 Clustering

In the first estimation step we recover the firm classes using a grouped fixed-effects approach (Bonhomme and Manresa, 2012), and use a k-means clustering algorithm for estimation. Consider the period-1 log wage distribution. Based on (14) the estimator solves:

$$\min_{f(1),...,f(J),H_1,...,H_L} \sum_{i=1}^{N} \int \left(\mathbf{1}\{Y_{i1} \leq y\} - H_{f(j_{i1})}(y)\right)^2 d\mu(y), \tag{15}$$

where the minimization is with respect to all possible groupings of firms into L groups, in addition to class-specific cdfs, and μ is a discrete or continuous measure.

Note that (15) is equivalent to the following weighted k-means problem:

$$\min_{f(1),...,f(J),H_1,...,H_L} \sum_{j=1}^{J} n_j \int \left(\widehat{F}_j(y) - H_{f(j)}(y)\right)^2 d\mu(y), \tag{16}$$

where $\widehat{F}_j$ denotes the empirical cdf of log wages in firm j, and n_j is the number of workers in firm j (both in period 1).

Verifying the assumptions of Theorems 1 and 2 in Bonhomme and Manresa (2012), one can show two large-sample properties of this estimator in an environment where firms are clustered in the population. First, the estimated firm classes, $\widehat{f}(j)$, converge uniformly to the population ones. Second, the asymptotic distribution of the class-ℓ cdf, $\widehat{H}_\ell$, coincides with that of the empirical cdf of wages in the population class ℓ (that is, the true one). In practice this means that, under suitable conditions, one can treat the estimated firm classes as data when computing standard errors of estimators based on them. In Proposition 6 below we prove a pointwise result for $\widehat{H}_\ell$, but the equivalence also holds uniformly in y.

For simplicity we take the measure μ to be discrete on $\{y_1, ..., y_D\}$. We will use the following assumptions, where $f^0(j)$ denote the population classes and H^0_ℓ denote the population class-specific cdfs, and where we denote $\|H\|^2 = \sum_{d=1}^{D} H(y_d)^2$.

Assumption 6. *(clustering)*

i) Y_{i1} *are independent across workers and firms.*

ii) For all $\ell \in \{1, ..., L\}$, $\text{plim}_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \mathbf{1}\{f^0(j) = \ell\} > 0$.

iii) For all $\ell \neq \ell'$ *in* $\{1, ..., L\}$, $\|H^0_\ell - H^0_{\ell'}\| > 0$.

iv) There exist $\underline{c} > 0, \overline{c} > 0$ *and* $\delta > 0$ *such that, for all* j, $\underline{c}n \leq n_j \leq \overline{c}n$, *and* $J/n^\delta \to 0$ *as* n *tends to infinity.*

Assumption 6 *i)* could be relaxed to allow for weak dependence both across and within firms, in the spirit of the analysis of Bonhomme and Manresa (2012) who focused on a time-series context. Assumptions 6 *ii)* and *iii)* require that the clusters be large and well-separated in the population. Assumption 6 *iv)* allows for asymptotic sequences where the number of workers per firm grows polynomially more slowly than the number of firms.

Proposition 2. *Let Assumption 6 hold. Then:*

i) $\Pr\left(\exists j \in \{1, ..., J\}, \widehat{f}(j) \neq f^0(j)\right) = o(1)$.

ii) For all y, $\sqrt{N_\ell}\left(\widehat{H}_\ell(y) - H^0_\ell(y)\right) \xrightarrow{d} \mathcal{N}\left(0, H^0_\ell(y)\left(1 - H^0_\ell(y)\right)\right)$, *where* N_ℓ *is the number of workers in class-*ℓ *firms; that is:* $N_\ell = \sum_{i=1}^{N} \mathbf{1}\left\{f^0(j_{i1}) = \ell\right\}$.

Proof. See Appendix A. ∎

As we pointed out in Section 3, adding information from job movers could be helpful, particularly when the separation condition in Assumption 6 *iii*) fails to hold. An alternative algorithm to estimate the firm classes is to solve:

$$\min_{f(1),...,f(J),G_{11},...,G_{LL}} \sum_{i=1}^{\widetilde{N}} \int \int \left(\mathbf{1}\{Y_{i1} \leq y\}\mathbf{1}\{Y_{i2} \leq y'\} - G_{f(j_{i1}),f(j_{i2})}(y,y')\right)^2 d\mu(y,y'), \tag{17}$$

for a bivariate measure μ, where the first $\widetilde{N} \leq N$ individuals are the job movers. A third possibility, which preserves the connection to the k-means algorithm, is to solve:

$$\min_{f(1),...,f(J),G_1,...,G_L} \sum_{i=1}^{\widetilde{N}} \int \int \left(\mathbf{1}\{Y_{i1} \leq y\}\mathbf{1}\{Y_{i2} \leq y'\} - G_{f(j_{i1})}(y,y')\right)^2 d\mu(y,y'), \tag{18}$$

for $G_1, ..., G_L$ a set of bivariate cdfs.

Finally, note that here we are clustering firm-specific cdfs. One could also base the clustering estimation on firm means, variances, or other moments of the firm-specific wage distributions. In addition, under similar assumptions one could use additional firm-level measurements in the clustering stage, such as output, profit or value added.

4.2 Finite mixtures

Once the firm classes $\widehat{f}(j)$ have been computed, in a second step we estimate $F_{k\ell}$ and $p_k(\ell,\ell')$. To do so, one possibility, which we use in the empirical analysis below, is to use a parametric model such as a Gaussian mixture model with (k,ℓ)-specific wage means and variances. We also experimented with a specification where $F_{k\ell}$ follows a mixture-of-normals with (k,ℓ)-specific parameters. The log-likelihood conditional on estimated firm classes takes the form:

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

where $f_{k\ell}(y;\theta)$ is a parametric wage density indexed by θ, and the proportions $p_k(\ell,\ell')$ are treated as parameters. We use the EM algorithm (Dempster, Laird, and Rubin, 1977) for estimation.

Several methods have recently been proposed to estimate finite mixture models while treating the type-conditional cdfs nonparametrically. See for example Bonhomme, Jochmans, and Robin (2014) and Levine, Hunter, and Chauveau (2011). These methods could also be used in the present context.

Finally, in the third estimation step, given estimates of the firm classes and the cdfs $F_{k\ell}$ we estimate the type proportions $\pi_k(\ell)$ based on another, simpler, finite mixture problem. The log-likelihood conditional on estimated firm classes and distributions is then, in period 1:

$$\sum_{i=1}^{N}\sum_{\ell=1}^{L}\mathbf{1}\{\widehat{f}(j_{i1})=\ell\}\ln\left(\sum_{k=1}^{K}\pi_k(\ell)f_{k\ell}(Y_{i1};\widehat{\theta})\right). \tag{19}$$

In practice we use a second EM algorithm to maximize (19). Note that the proportions could also be estimated using a least squares regression, as can be seen from the identification arguments in Section 3.

4.3 Adding worker covariates

Our aim is to document sorting and complementarity between firm and worker characteristics. Some of these characteristics (such as worker age or education, or firm size) are typically available in matched data. Here we outline how we incorporate these observed covariates in estimation.

The log wages we consider in the application are net of quarterly dummies, and year indicators interacted with cohort dummies, within education groups. We remove these time effects in order to make the two periods more comparable. An alternative approach would be to estimate the non-stationary model analyzed in Section 7. We estimate the firm classes by clustering log wage distributions net of time effects. The estimated classes thus reflect differences in firm and worker observables as well as unobservables. In the grouped fixed-effects approach, the classes can be unrestrictedly correlated to firm-level covariates and to composition in terms of worker covariates.

We estimate the cdfs $F_{k\ell}$ on log wages net of time effects. In this specification, the effects of worker types reflect both observable and unobservable attributes. In order to separate the effects of observables from the effects of unobservables, we estimate the proportions $\pi_k(\ell,x)$ by allowing them to depend on time-invariant worker covariates, X_i, such as education or age in the first period. The conditional period-1 log-likelihood is then:

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

where $\pi_k(\ell,x)$ are treated as parameters.

This specification allows us to distinguish sorting in terms of x from sorting in terms of unobservables. For example, for all (k,ℓ) we can write:

$$\pi_k(\ell) \;=\; \sum_x p_x\pi_k(\ell,x) \;+\; \sum_x \left(p_x(\ell)-p_x\right)\pi_k(\ell,x), \tag{20}$$

where $p_x = \Pr(X_i = x)$, and $p_x(\ell) = \Pr(X_i = x | f(j_{i1}) = \ell)$. The first term on the right-hand side of (20), say $\widetilde{\pi}_k(\ell)$, represents the type proportion in a counterfactual economy where covariates x are equally distributed across firm classes. Hence the two terms on the right-hand side of (20) reflect the contribution of unobservables and observables, respectively, to differences in worker type composition across firm classes.

4.4 Interpreting the estimates

With estimates of type proportions and conditional wage distributions at hand, one can perform distributional decomposition exercises. For example, shutting down sorting (thus setting $\pi_k(\ell)$ independent of ℓ) one can generate the wage distribution that would prevail under random assignment of workers to firms:

$$\widetilde{H}(y) = \sum_{\ell=1}^{L} q_\ell \sum_{k=1}^{K} \overline{\pi}_k F_{k\ell}(y), \tag{21}$$

where $\overline{\pi}_k = \Pr(\omega(i) = k)$, and $q_\ell = \Pr(f(j_{i1}) = \ell)$.

In order to relate our results with those obtained using the AKM method, we will also report estimates of variance decompositions performed on data simulated from our model. These estimates are obtained by projecting log wages on worker type dummies and firm class dummies in an additive manner:

$$Y_{i1} = \alpha \left[\omega^0(i)\right] + \psi \left[f^0(j_{i1})\right] + \varepsilon_{i1}, \tag{22}$$

where $\omega^0(i)$ and $f^0(j_{i1})$ are the population types and classes, respectively. From (22) we can compute the variances of worker and firm effects, and the covariance between the two, as in AKM. In addition, the R-squared in (22) can be compared to the one in a fully saturated regression that allows for unrestricted interactions between $\omega^0(i)$ and $f^0(j_{i1})$, thus shedding light on the quantitative importance of interaction effects.

5 Data patterns (preliminary)

5.1 Description of the data

We use a match of four different databases from Friedrich, Laun, Meghir, and Pistaferri (2014) covering the entire working age population in Sweden between 1997 and 2006. Out of the four data sources the Swedish data registry, ANST (which is part of the register-based labor market

statistics at Statistics Sweden) links information about individuals, their employment, and their employer. This database is collected yearly from the firm's statement of income for their employees. The other databases provide additional information on workers and firms as well as unemployment status of workers. These are LOUISE (or LINDA), which contains information on the workers; SBS, which provides accounting data and balance sheet information for all non-financial corporations in Sweden; and the Unemployment Register, which contains spells of unemployment registered at the Public Employment Service.

Following Friedrich, Laun, Meghir, and Pistaferri (2014) we perform a first sample selection by dropping all financial corporations and some sectors such as fishery and agriculture, education, health and social work.

Employment spells and wages. The RAMS dataset allows to construct individual employment spells since it provides the first and last remunerated month for each employment, as well as firm and plant identifier. In this version of the paper we only make use of firm identifiers. Moreover, RAMS provides yearly pre-tax earnings for each spell. We use these earnings as our measure of wages since we do not observe hours worked. With some abuse of terminology we will refer to these full-year pre-tax labor earnings as "wages". We discard employment spells of workers with more than one employment spell in a given year. In particular, if a worker has multiple employers, only the job where he is earning the highest wage is recorded, irrespective of the number of months he worked in each firm.

In this version of the paper we remove age and time effects by regressing log wages on interactions of time dummies with cohort dummies, within education groups. This implies that the mean log wage is constant over time within education and cohort, equal to the value corresponding to the first period in our final sample (2002, see below for a description). We do so in order to remove the effects of life cycle variation and macroeconomic shocks. Another approach would be to estimate a model with time-varying wage distributions, which we analyze in Section 7.

Selection of the samples. In the two-period version of our model we rely on information from the bivariate wage distributions of job movers, and the univariate cross-sectional wage distributions, in different firms. We now describe the two different samples that we use to estimate our model: Samples 1 and 2.

In order to construct Sample 1 we restrict our attention to workers employed in years 2002 and 2004. These two years correspond to period 1 and 2 in the model. We also restrict the

sample to males in order to avoid dealing with gender differences in labor supply, since we do not have information on hours worked. Finally, we select workers fully employed in a firm in both 2002 and 2004, in order to avoid selecting individuals with low labor market attachment.

Table 1: Data description

	all	Sample 1	Sample 2
time spell	1997 - 2008	2002 - 2004	2002 - 2004
number of worker IDs	2,155,707	974,860	20,072
number of firm IDs	87,560	59,325	12,976
firm $\geq$ 10 workers	73,400	33,035	-
firm $\geq$ 50 workers	35,062	6,829	-
firm $\geq$ 5 movers	58,989	3,718	1,400
firm $\geq$ 10 movers	45,086	1,822	552
firm $\geq$ 20 movers	29,651	933	230

Notes: Sample 2 contains the job movers from Sample 1, where job mobility has occurred between December 2002 and January 2004.

To construct Sample 2 we select the job movers of Sample 1, that is, workers whose employer firm identifier is different in 2002 and 2004. Moreover, in order to avoid considering job changes that are unrelated to job mobility we discard workers whose firm identifier is not present in either of the two periods. A firm may appear in only one period because of a merger or acquisition, entry or exit of a firm, or a re-definition of the firm identifier over time. We choose not to interpret these changes as job moves as they do not map naturally to our model.

Table 1 shows a few descriptive statistics for the full 1997-2006 sample, Sample 1, and Sample 2. The median firm size, as reported by the employer, is 10 workers for firms in Sample 1, and 27 workers for firms that appear in Sample 2. The median number of full-year employed males per firm is only 4 workers for firms in Sample 1. However, when focusing on firms that appear in Sample 2 (that is, which have more than one job mover) the median number of full-year employed males per firm is 13 workers.

5.2 Descriptive evidence

As outlined in Section 4, we estimate firm classes using a weighted k-means algorithm. In this version of the paper we cluster firms based on their cross-sectional wage distributions in

2002; see equation (15). Cumulative distribution functions of log wages are evaluated at 40 percentiles of the overall 2002 wage distribution. We allow for $L = 10$ firm classes. The results in Table 2 correspond to the minimum value of the objective function among 1000 randomly generated starting values. In the table we have ordered firm classes according to mean log wage in each class (although the ordering of the classes is arbitrary in our setting).

Table 2: Descriptive statistics on estimated firm classes

firm cluster:	1	2	3	4	5	6	7	8	9	10	all
number of workers	21,662	62,929	110,792	114,324	100,080	78,837	137,971	85,806	58,728	27,023	798,152
number of firms	6,487	7,972	7,804	6,494	4,663	3,748	4,209	3,984	3,157	2,812	51,330
% HS dropout	28.9	28	26.6	26.9	23.7	21.1	18.9	12.2	5.31	3	20.7
% HS grade	59.7	62.5	62.6	62.5	61.7	57.8	58.6	47.2	32.9	23.9	56.1
% some college	11.4	9.42	10.7	10.7	14.6	21.2	22.5	40.5	61.8	73.1	23.3
% workers younger than 30	25	20.9	20.5	18.4	16.4	18.4	14.3	15.3	16	14.8	17.5
% workers between 31 and 50	52.7	52.2	53.2	54.1	55.4	54.6	56	57	59.1	63.6	55.3
% workers older than 51	22.3	26.9	26.2	27.5	28.2	27.1	29.7	27.7	24.9	21.6	27.2
mean log wages	9.6	9.87	9.99	10.1	10.1	10.1	10.2	10.4	10.5	10.8	10.2
variance of log wages	0.15	0.0841	0.0934	0.0732	0.0699	0.141	0.0918	0.114	0.116	0.177	0.148
between firm variance of log wages	0.0576	0.00614	0.0039	0.00185	0.0016	0.0056	0.00184	0.00367	0.00456	0.039	0.0544
mean of log value added per worker	12.4	12.5	12.7	12.7	12.8	12.8	12.9	13	13	13.2	12.7
variance of log value added per worker	0.202	0.163	0.155	0.141	0.175	0.255	0.249	0.304	0.398	0.594	0.52
median number of workers per firm	2	4	6	5	5	6	5	5	4	3	4

Notes: Sample 1 in 2002. All workers are males, employed during the full year 2002. "HS" is high school.

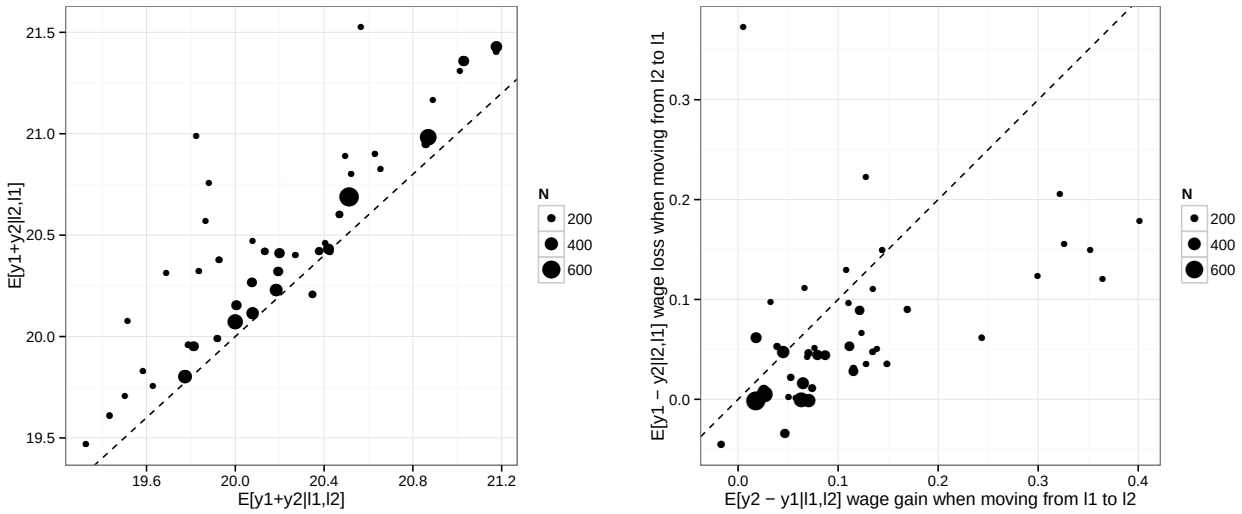
Table 2 shows that firm classes capture substantial heterogeneity between firms. The 10 classes seem to capture wage heterogeneity well. In particular, the between-firm-class log wage variance is 90% of the overall between-firm variance. There are also substantial differences between classes in terms of the level of education of workers: while the lower classes show high percentages of high school dropouts or high school graduates and low percentages of workers with some college, the higher classes show the opposite pattern. In terms of workers' age, we observe that lower classes tend to have higher percentages of young workers (less than 30 years old) and lower percentages of workers between 30 and 50, while higher classes have more workers between 30 and 50. This relationship broadly reflects the life cycle pattern of wages in these data. Workers above 51 years old are more evenly distributed across firm classes.

We also observe that the number of sample workers per firm follows an inverse U-shape in firm class, with the highest and lowest classes tending to have smaller firms. Note that median firm sizes are higher than these numbers, as here we focus on full-year employed males. Moreover, as we discussed following Table 1, most of the smaller firms have no job movers, so they will not contribute to the second estimation step. Lastly, Table 2 shows that log valued

added per worker is monotone in firm class, suggesting that the ordering based on mean log wages agrees with an ordering in terms of productivity. However, it is worth noting that the classes explain only 19% of the between-firm variance in log valued added per worker, suggesting that productivity differences within classes are substantial.

Wages of job movers. Job movers are central to our identification strategy. Motivated by the discussion in Subsection 3.1, in Figure 2 we report two sets of results. On the left graph we plot the relationship between the sum of log wages in the two periods for workers with opposite movements between firm classes ℓ_1 and ℓ_2, where $\ell_1 < \ell_2$. The size of each dot represents the number of workers in the sample making the transition in either direction. The graph clearly shows that average log wages are higher for workers moving from ℓ_2 to ℓ_1 than for workers moving from ℓ_1 to ℓ_2. In words, the wages of job movers are more closely related to the firm they come from, than to the firm they move to. This suggests that worker type composition differs between the two types of transitions, and is suggestive of the presence of the sort of asymmetry that we need for identification.

Figure 2: Wages of job movers



Notes: Wages of job movers between firm classes ℓ_1 and ℓ_2, where $\ell_1 < \ell_2$. The left graph shows the sum of log wages in periods 1 and 2 (that is, 2002 and 2004) when moving from ℓ_1 to ℓ_2 (x-axis) versus those when moving from ℓ_2 to ℓ_1 (y-axis). The right graph shows log wage gains (x-axis) and losses (y-axis) corresponding to the same transitions between firm classes.

The right graph in Figure 2 plots the log wage gains (x-axis) and losses (y-axis) corresponding to opposite movements between firm classes. As we discussed in Subsection 3.1, under

additivity of firm and worker heterogeneity all the dots would lie on the 45 degree line. The graph shows that, on average, the dots lie below the line, suggesting that job mobility tends to be associated with wage gains. Note that time effects (interacted with cohort and education dummies) were taken out from log wages in the full sample, while here we are considering a subsample of job movers. Moreover, the right graph in Figure 2 does show some pairs of firm classes for which the dots lie far from the 45 degree lines. This suggests some departure from additivity, and motivates the analysis of the next section, where we present estimates of sorting and wage returns that do not rely on additivity.

6 Empirical results (preliminary)

In this section we present estimates of conditional wage distributions and worker/firm sorting. We then report several decomposition exercises on both our data and simulated samples.

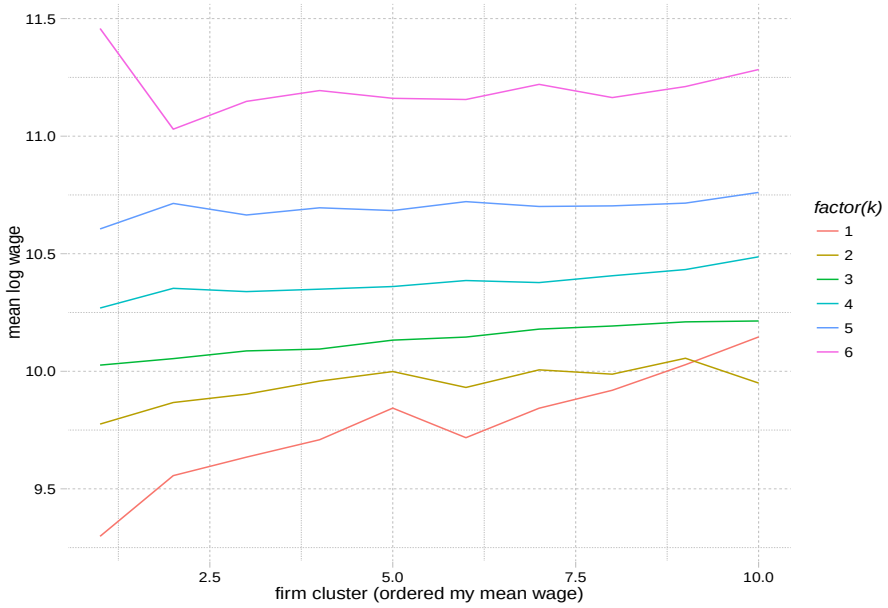
6.1 Estimates of wage distributions

Given the estimated firm classes, we run an EM algorithm on job movers (Sample 2) to recover the cdf estimates $\widehat{F}_{k\ell}$ and the type proportions among job movers $\widehat{p}_k(\ell, \ell')$. Conditional distributions are Gaussian, with class-and-type-specific means and variances. Finding the global maximum of the likelihood can be challenging in mixture models. We used several strategies to pick sensible starting values. We first took as starting values the parameter estimates from a simplified model where means and variances are assumed not to depend on firm classes. We observed that this technique tended to give good starting values. In addition we also considered a large set of starting values drawn at random.

Figure 3 shows the mean estimates, for $L = 10$ firm classes and $K = 6$ worker types. On the x-axis, firm classes are ordered by mean log wages. The results show clear evidence of worker heterogeneity. They also show some variation with firm class. However, the highest two worker types (in terms of mean log wages) experience relatively little wage variation when working in different firms. In contrast, the bottom four types show a clear positive gradient with firm class, especially so for the lowest type. This finding points towards larger complementarity between low type workers and firm classes.

These preliminary results do not seem consistent with additive models of worker and firm unobservables, such as AKM. These models would predict equal mean log wage gains across firms for different types of workers, which is not consistent with the presence of higher complementarity in the bottom part of the worker distribution. At the same time, note that there

Figure 3: Estimated means of log wages by worker type and firm class



Notes: The graph plots the mean of $\widehat{F}_{k\ell}$*. The* $L = 10$ *firm classes (on the x-axis) are ordered by mean log wage. The* $K = 6$ *worker types correspond to the 6 different curves.*

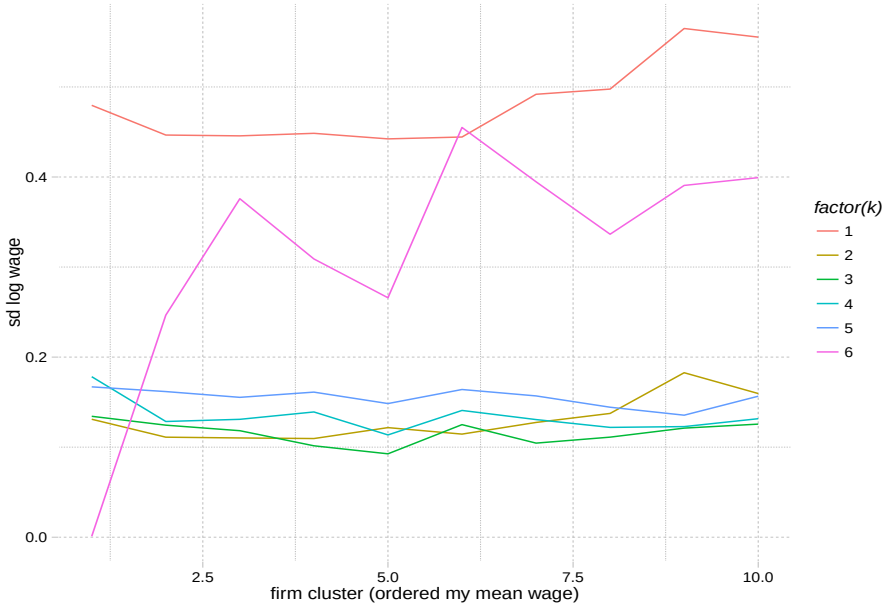
are no standard errors on the graph. In particular, the decrease for type 1 workers between the first two firm classes is likely to be statistically insignificant, because the estimated probability (reported below) of the $(k = 1, \ell = 1)$ combination is very small. More generally, Figure 3 alone does not allow to determine whether the evidence of non-additivity is quantitatively relevant. We will return to this question in the next subsection.

Figure 4 shows the estimated standard deviations by worker type and firm class. The lowest worker type (in terms of mean log wages) has the largest standard deviation. The highest worker type also has a large standard deviation. This suggests the presence of considerable heterogeneity within these worker types. A higher variance might also reflect the presence of higher wage uncertainty for these workers. Lastly, the four middle worker types face much smaller standard deviations.

6.2 Estimates of worker/firm sorting

Given the estimated wage distributions, we then run a second EM algorithm on the 2002 cross-section (the first period in Sample 1) in order to recover the worker type proportions in the cross-section $\widehat{\pi}_k(\ell, x)$; see (19). Unlike the first one, this second EM algorithm is not sensitive

Figure 4: Estimated standard deviations of log wages by worker type and firm class



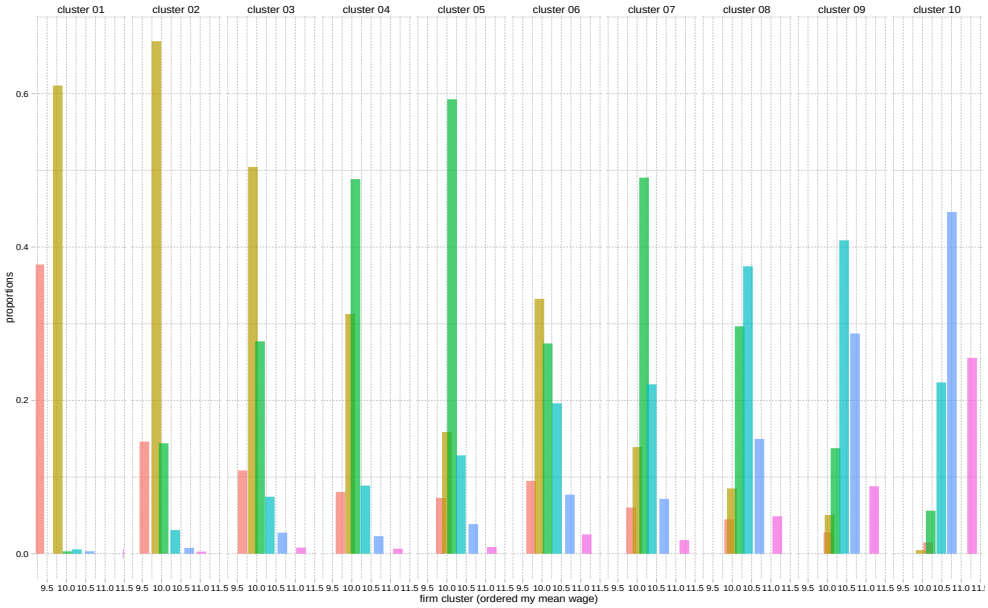
Notes: The graph plots the standard deviation of $\widehat{F}_{k\ell}$*. The* $L = 10$ *firm classes (on the x-axis) are ordered by mean log wage. The* $K = 6$ *worker types correspond to the 6 different curves.*

to the choice of starting values as the log-likelihood function is concave. We allow the type proportions to depend on 3 age categories (terciles of the age distribution) and 3 education categories (high school dropouts, high school graduates, some college); hence there are 9 cells x.

Figure 5 shows the proportions of worker types in each firm class $\widehat{\pi}_k(\ell)$. Figure 6 shows the same proportions, net of differences in education and age: $\widetilde{\pi}_k(\ell) = \sum_x \widehat{p}_x \widehat{\pi}_k(\ell, x)$, where $\widehat{p}_x$ is the number of workers in cell x; see (20). Figure 5 thus shows how much sorting there is in terms of worker observables and unobservables, while Figure 6 reflects sorting due to unobservable characteristics only. On each graph the x-axis corresponds to the mean of $\widehat{F}_{k\ell}$. The results show strong evidence of cross-sectional sorting, mostly driven by sorting on unobservables. The types of workers differ markedly across firm classes. For example, the lowest-class firms (in terms of mean log wage) employ mostly the bottom two worker types, while the highest-class firms employ mostly the high worker types.

Lastly, Figure 7 plots the estimated probability of moving to a higher firm class for different worker types, where classes are ordered with respect to mean log wages. The graph shows, for each worker type, the sum of the probabilities of making a transition from ℓ to $\ell' > \ell$ conditional on moving to another firm, see (8), averaged over firm classes. This probability is

Figure 5: Estimated proportions of worker types by firm class



Notes: The graph plots the type proportions $\widehat{\pi}_k(\ell)$. Each plot corresponds a different firm class. Average log wages by (k, ℓ) combinations are on the x-axis, proportions are on the y-axis.

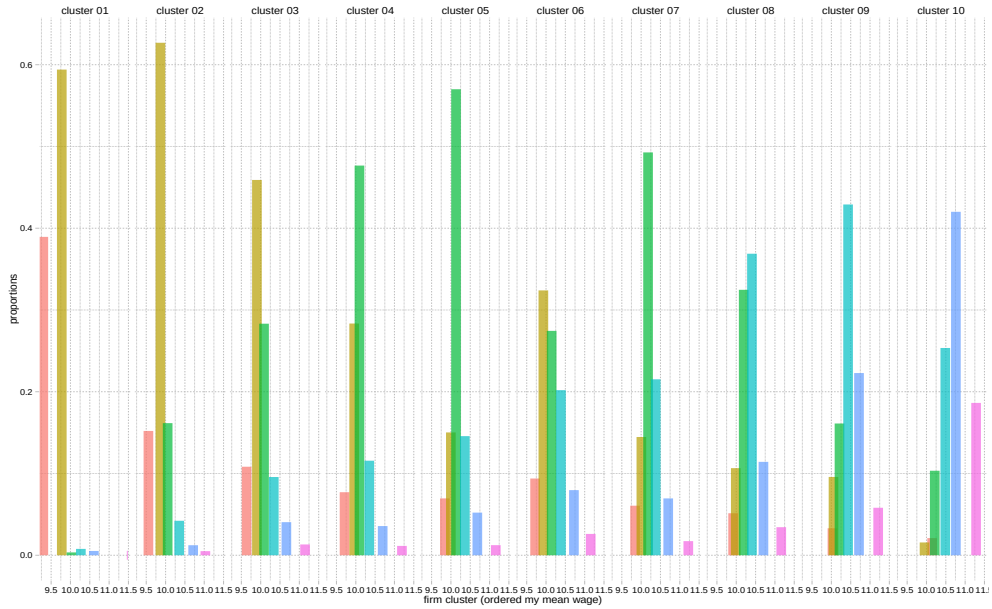
increasing in worker type, ranging from 38% for the lowest type to 60% for the highest worker type. We interpret this evidence as reflecting the presence of dynamic sorting, in addition to the cross-sectional sorting documented in Figure 5.

6.3 Decomposition exercises

In this subsection we report the results of decomposition exercises on simulated data and on the Swedish data. We start by describing the simulation method.

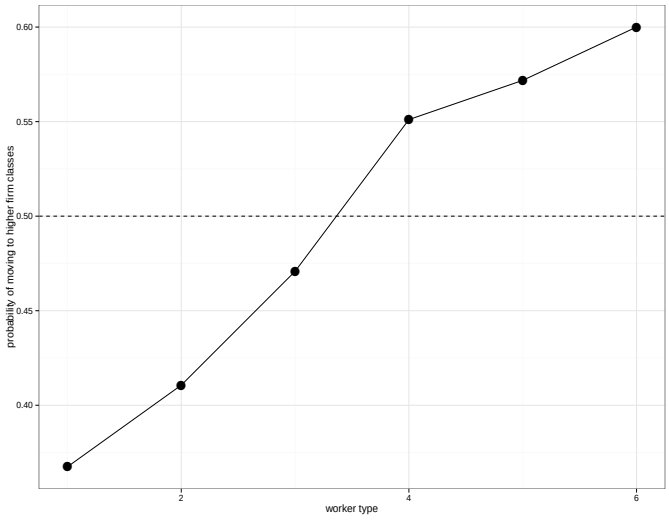
Model simulation and fit. To simulate from the model we use the estimated parameters reported in the previous subsections, and we condition on the firm identifiers and job movements observed in the data. The model allows to simulate the wages of job movers in 2002 and 2004. Consider first the 2002 wage. We get from the data the length of the job spell that ended after December 2002, and we draw the worker type k from the corresponding $p_k(\ell, \ell')$ (where ℓ and ℓ' are the estimated classes of the firms where the worker was employed in 2002 and 2004, respectively), and the log wages from the corresponding cdf $F_{k\ell}$. In addition, we add serial correlation between different draws from that distribution. The correlation coefficient

Figure 6: Estimated proportions of worker types by firm class (no covariates)



Notes: The graph plots the type proportions, net of differences in education and age x: $\widetilde{\pi}_k(\ell) = \sum_x \widehat{p}_x \widehat{\pi}_k(\ell, x)$, where $\widehat{p}_x$ is the number of workers in cell x; see (20). Each plot corresponds a different firm class. Average log wages by (k, ℓ) combinations are on the x-axis, proportions are on the y-axis.

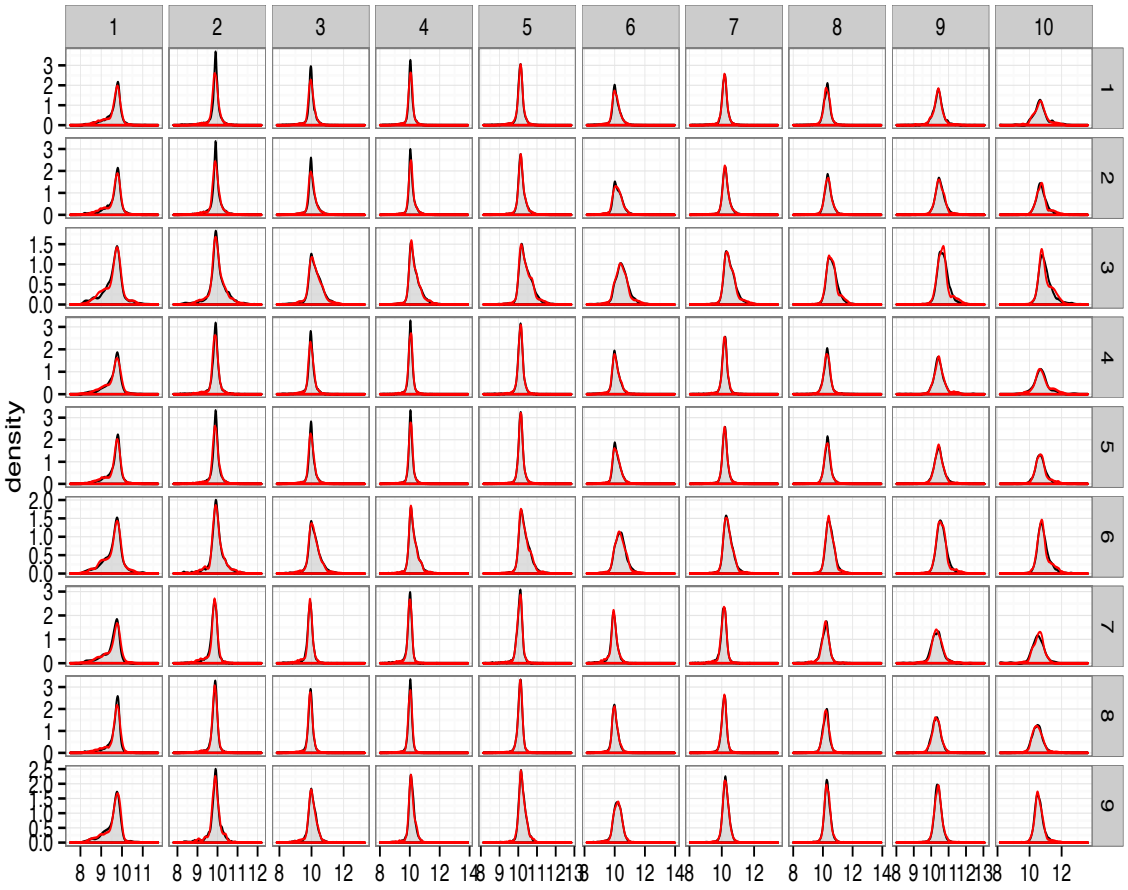
Figure 7: Probability of moving to a higher firm class



Notes: This graph plots the estimated probability of moving to a higher firm class (y-axis), by worker type (x-axis), averaged over classes; that is: $\sum_{\ell' > \ell} \Pr\left[f(j_{i2}) = \ell' \mid f(j_{i1}) = \ell, \omega(i) = k, m_i = 1\right]$, see (8).

is chosen in order to approximate the within-job log wage autocorrelation in the data. We proceed similarly for spells that started before January 2004, and for non job movers. Though not necessary for using our model and checking its fit, for example, simulating full employment spells will allow us to compare our results with those obtained using the AKM methodology.

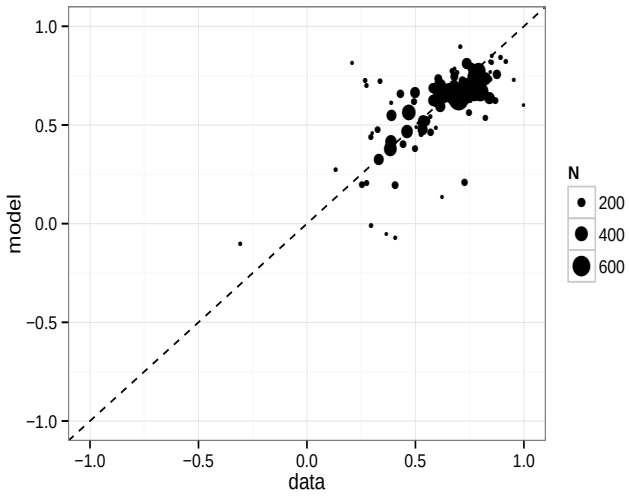
Figure 8: Fit of log wage densities



Notes: Marginal densities of log wages for each x cell (in rows) and firm class (in columns). Sample 1, 2002. The red line is the model, the shaded area is from the data.

Figure 8 shows the model fit for the densities of log wages, by covariates cells x and firm classes. The fit is excellent. Figure 9 shows the model fit for the correlation between 2002 and 2004 log wages, for the job movers in Sample 2, by pairs of firm classes. Although there is some discrepancy between the correlations computed in the data and those generated by the model, the fit is generally good.

Figure 9: Fit of log wage correlations



Notes: Log wage correlations $\text{Corr}(Y_{i1}, Y_{i2}|\ell_1, \ell_2)$*, for job movers, by pairs of firm classes. Sample 2. In the data (x-axis) and in the simulated data (y-axis).*

Variance decompositions. Using the simulated data, we then estimate a linear regression of log wages on worker type dummies and firm class dummies; see (22). We then perform a variance decomposition. The first row in Table 3 shows that worker heterogeneity explains substantially more than firm heterogeneity according to our estimates (referred to as "BLM" in the table). Indeed, the variance of firm class coefficients is only 6% of that of worker type coefficients. Moreover, the correlation between the two sets of coefficients is 46%, which is in line with the strong evidence of sorting documented in Figure 5.

The R-squared of the linear regression is 74.8%. It is of interest to compare it to the R-squared of a saturated regression with full interactions between worker types and firm classes, which is 75.6%. The small difference between the two suggests that an additive specification provides a good approximation to the conditional mean of log wages.

In the second row of Table 3 we report the results of an AKM regression on the Swedish data. We follow the literature and use raw log wages but include covariates in the regression. We add a cubic in age, interacted with the three levels of education. As in Card, Heining, and Kline (2013) we first estimate individual and firm fixed effects on the subsample of job movers (including their entire job spells, before and after the move). This requires to construct the largest connected set in our sample, which represents 9,079 firms and 17,595 movers (that is, 88% of Sample 2). Given the estimated firm fixed effects, we then estimate individual fixed

effects for the non-movers.

The results of the AKM estimation are very different from the ones using our model. The variance of firm fixed-effects is one third of that of worker fixed-effects. Moreover, the correlation between firm and worker fixed-effects is negative, equal to -26%. These orders of magnitude are within the range of the empirical results obtained using AKM in the literature.

In order to better understand these differences, we then repeat the above exercises on a sample simulated from our model. The third row of Table 3 shows that our method recovers the parameters quite well, which is consistent with the good model fit reported above. In contrast, the AKM results reported on the fourth row of the table show a strong negative correlation (-35%), in a data generating process where the true correlation is positive and substantial (46%). In addition, AKM strongly overestimates the contribution of firm heterogeneity to the variance of log wages.

Table 3: Variance decompositions on Swedish data and simulated data

	min spell	rep	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$
			Data			
BLM			0.7766	0.0473	0.1762	0.4598
AKM			0.9813	0.3014	-0.2826	-0.2599
			Simulated from BLM			
BLM	1	1	0.7669	0.0466	0.1866	0.4934
AKM	1	1	1.0879	0.3447	-0.4326	-0.3532
			Simulated from BLM without limited mobility			
AKM	4	1	0.8948	0.1602	-0.055	-0.0727
AKM	4	10	0.7816	0.053	0.1654	0.4064

Notes: Real and simulated data. α is the worker (or type) fixed-effect, ψ is the firm (or class) fixed-effect. "BLM" is our approach. "min spell" is the minimum length of employment spells. "rep" is the number of job movers per firm, relative to the original dataset.

These results suggest a poor performance of AKM on these data, while in contrast our method seems to perform well. A possible explanation for the AKM bias is given by Andrews, Gill, Schank, and Upward (2008): in the presence of few job movers per firm, the correlation estimated using the AKM method may be downwardly biased, even negative when the true correlation is positive. This motivates the final exercise, which we report in rows 5 and 6 of

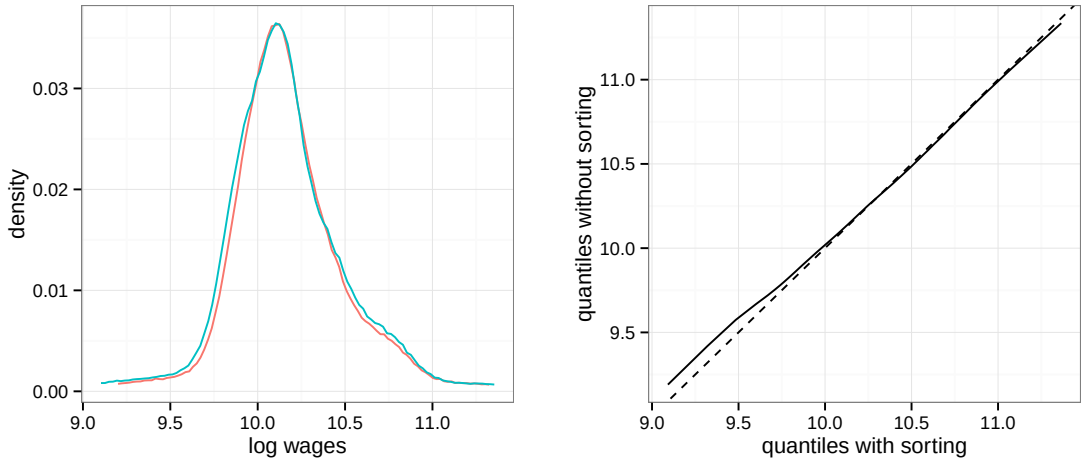
Table 3. When we artificially increase the number of movers per firm (10 times) and increase the length of job spells (we impose of minimum of 4 years per spell), AKM recovers the correlation and the percentage of variance explained by the firms rather well. Note that this is despite the fact that the data generating process is non-additive.

Overall, these exercises suggest that, while we uncover evidence of non-additivity between worker and firm heterogeneity, an additive specification may not be quantitatively misleading on these data. At the same time, the AKM method seems not to properly recover the amount of sorting, and more generally the contributions of worker and firm heterogeneity to wage dispersion, unless the data contain a large number of job movers per firm and long employment spells. In contrast, our method seems well suited to deal with short panels with relatively few job movements.

Sorting and wage inequality. In order to assess the impact of worker/firm sorting on wage inequality, we next plot a counterfactual log wage density, in 2002, assuming random assignment of workers to firms, versus the density estimate reproduced by the model. The counterfactual density corresponds to the cdf in (21). The left graph in Figure 10 shows that the two densities are close to each other. This reflects the fact that, according to our estimates, while sorting is substantial, the firm component explains relatively little of the variance of log wages. At the same time, the counterfactual wage distribution without sorting is less dispersed than the empirical one: the variance is .127 without sorting, versus .144 with sorting (88%). This suggests that sorting does explain a non-negligible part of wage dispersion. On the right graph in Figure 10 we compare the quantiles of log wages as reproduced by the model, with the quantiles in the counterfactual economy with no sorting. We see that workers at the bottom of the distribution experience higher wages in the absence of sorting, while the differences at the top of the distribution are small. This is consistent with the evidence of higher complementarity between firms and low wage workers (Figure 3).

Simulations from a theoretical sorting model. Lastly, in Appendix B we evaluate the performance of our method to recover the contributions of worker and firm heterogeneity to wage dispersion, when the data generating process follows a calibrated version of the model of Shimer and Smith (2000) with on-the-job search (see Moscarini, 2005; Shimer, 2006; Flinn and Mabli, 2008 for related contracting environments). We analyze two situations: with positive assortative matching (PAM) and with negative assortative matching (NAM). We show that our approach recovers the contributions of firms and workers well, even when the model has a large

Figure 10: Sorting and wage dispersion



Notes: The left graph plots the log wage densities, Sample 1, 2002. Blue is simulated from the model, red is simulated from the model without sorting. The right graph plots the quantiles of log wages in the simulation without sorting (y-axis) against the quantiles in the simulation with sorting (x-axis), with the 45 degree line in dashed.

number of latent types and classes ($K = 50$, $L = 50$) but we use the same number of types and classes as in the empirical analysis ($K = 6$, $L = 10$). In addition, we show that our approach estimates sorting patterns rather well, even under NAM.

Given these results, a natural question is whether existing theoretical sorting models can account for the empirical evidence that our method has uncovered on the Swedish data. The combination of strong sorting of high wage workers to high wage firms, and complementarity between low wage workers and high wage firms, represents a challenge for models built on Becker (1974), according to which positive sorting goes hand in hand with positive complementarity between worker and firm heterogeneity. Although this exceeds the scope of this paper, building structural models that can match the empirical evidence is an important research question.

7 Extensions

In this section we describe several extensions of the main analysis.

7.1 Non-stationary model

In the model of Section 2, we assumed that wage distributions in different periods, $F_{k\ell}$, were time-invariant. Relaxing Assumption 3 may be desirable for several reasons: to allow for calendar time effects such as macroeconomic shocks, and for unrestricted dependence of $F_{k\ell}$ on worker (or firm) age, as well as to allow for wages before and after a job move to be drawn from different distributions.

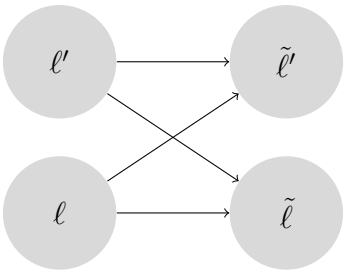
Allowing for non-stationary wage distributions, the bivariate log wage distribution of job movers between period 1 and 2 becomes:

$$\Pr\left[Y_{i1} \leq y\,,\, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] = \sum_{k=1}^{K} p_k^{12}\left(\ell, \ell'\right) F_{k\ell}^1(y) F_{k\ell'}^2(y'). \tag{23}$$

To show identification of $F_{k\ell}^1$, $F_{k\ell'}^2$, and $p_k^{12}(\ell, \ell')$ based on (23), we start with a definition, which we illustrate in Figure 11.

Definition 2. *An alternating cycle of length R is a pair of sequences of firm classes $(\ell_1, ..., \ell_R) \in \{1, ..., L\}^R$, with $\ell_{R+1} = \ell_1$, and $(\widetilde{\ell}_1, ..., \widetilde{\ell}_R) \in \{1, ..., L\}^R$, such that $p_k(\ell_r, \widetilde{\ell}_r) \neq 0$ and $p_k(\ell_{r+1}, \widetilde{\ell}_r) \neq 0$ for all $r \in \{1, ..., R\}$ and $k \in \{1, ..., K\}$.*

Figure 11: An alternating cycle containing ℓ and ℓ'



Assumption 7. *(non-stationary model)*
i) For any two firm classes $\ell \neq \ell'$ in $\{1, ..., L\}$, there exists an alternating cycle $(\ell_1, ..., \ell_R)$, $(\widetilde{\ell}_1, ..., \widetilde{\ell}_R)$, such that $\ell_1 = \ell$ and $\ell_r = \ell'$ for some r, and such that the scalars $a_1, ..., a_K$ are all distinct, where:

$$a_k := \frac{p_k^{12}(\ell_1, \widetilde{\ell}_1) p_k^{12}(\ell_2, \widetilde{\ell}_2) ... p_k^{12}(\ell_R, \widetilde{\ell}_R)}{p_k^{12}(\ell_2, \widetilde{\ell}_1) p_k^{12}(\ell_3, \widetilde{\ell}_2) ... p_k^{12}(\ell_1, \widetilde{\ell}_R)}.$$

In addition, for all ℓ, ℓ', possibly equal, there exists an alternating cycle $(\ell'_1, ..., \ell'_R)$, $(\widetilde{\ell}'_1, ..., \widetilde{\ell}'_R)$, such that $\ell'_1 = \ell$ and $\widetilde{\ell}'_r = \ell'$ for some r.

ii) For a suitable finite set of values for y and y', and for all $r \in \{1, ..., R\}$, the matrices:

$$A(\ell_r, \widetilde{\ell}_r) := \left\{ \Pr \left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell_r, f(j_{i2}) = \widetilde{\ell}_r, m_i = 1 \right] \right\}_{(y,y')} \text{ and}$$

$$A(\ell_r, \widetilde{\ell}_{r+1}) := \left\{ \Pr \left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell_r, f(j_{i2}) = \widetilde{\ell}_{r+1}, m_i = 1 \right] \right\}_{(y,y')}$$

have rank K.

The following result shows that $F^1_{k\ell}$, $F^2_{k\ell'}$ and $p^{12}_k(\ell, \ell')$ are identified. Identification of type proportions $\pi^1_k(\ell)$, and of firm classes $f(j)$, follows as in the main analysis.

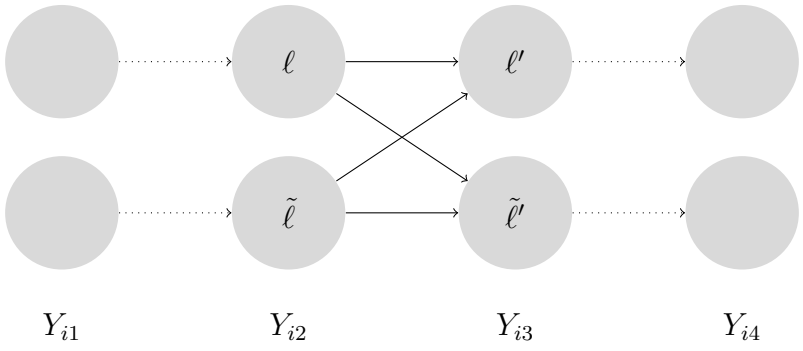
Theorem 2. *Consider the joint distribution of wages of job movers in the non-stationary model, see (23), and let Assumptions 1, 2 and 7 hold. Then, up to common relabelling of the k index, $F^1_{k\ell}$ and $F^2_{k\ell'}$ are identified for all (k, ℓ, ℓ'). Moreover, for all pairs (ℓ, ℓ') for which there are job moves from ℓ to ℓ' $p^{12}_k(\ell, \ell')$ is identified for all k, up to the same relabelling.*

Proof. See Appendix A. ∎

7.2 Endogenous mobility and serial correlation

Here we explain how to allow for endogenous mobility and serial correlation in an extension of the model of Section 2. For this, we rely on an additional period of observation, both before and after the job move. We consider four periods of employment, and denote the log wages as Y_{it} and the firm classes as $f(j_{it})$, for $t \in \{1, ..., 4\}$. We denote as $m_i = \mathbf{1}\{j_{i2} \neq j_{i3}\}$ the job mobility indicator between period 2 and 3. We make the following assumptions.

Figure 12: An alternating cycle in the model with endogenous mobility



Assumption 8.

i) Y_{i1} *is independent of* $(m_i, Y_{i3}, f(j_{i3}))$ *given* $(\omega(i), Y_{i2}, f(j_{i2}))$.

ii) Y_{i4} *is independent of* $(Y_{i2}, f(j_{i2}), Y_{i1})$ *given* $(\omega(i), Y_{i3}, f(j_{i3}), m_i = 1)$.

Assumption 8 has two parts. Part *i)* requires that mobility decisions only depend on the current firm class, the worker type, and the current wage, but not on the past wage, and that mobility outcomes of job movers are also conditionally independent of wages measured one period before the move. Part *ii)* requires that wages of job movers one period after the move only depend on the previous wage, the previous firm class, and the worker type, but not on the wages and firm class before the move.

Assumption 8 is more general than the assumptions we imposed on the model under strict exogeneity and serial independence. This extension allows to nest theoretical sorting models that were not consistent with the assumptions of Section 2. As an example, Assumption 8 is satisfied in the model of Shimer and Smith (2000) augmented with on-the-job search and match-specific heterogeneity, when the latter evolves as a first-order Markov process. It will also typically be satisfied in wage posting models based on Burdett and Mortensen (1998) that allow for two-sided heterogeneity and on-the-job wage dynamics.

Given Assumption 8 we have:

$$\begin{aligned}\Pr\left[Y_{i1} \leq y, Y_{i4} \leq y' | Y_{i2} = y_2, Y_{i3} = y_3, f(j_{i2}) = \ell, f(j_{i3}) = \ell', m_i = 1\right] \\ = \sum_{k=1}^{K} p_k\left(y_2, y_3, \ell, \ell'\right) G_{k\ell}(y|y_2) H_{k\ell'}(y'|y_3),\end{aligned} \tag{24}$$

where:

$$\begin{aligned} p_k\left(y_2, y_3, \ell, \ell'\right) &:= \Pr\left[\omega(i) = k | Y_{i2} = y_2, Y_{i3} = y_3, f(j_{i2}) = \ell, f(j_{i3}) = \ell', m_i = 1\right], \\ G_{k\ell}(y|y_2) &:= \Pr\left[Y_{i1} \leq y | \omega(i) = k, Y_{i2} = y_2, f(j_{i2}) = \ell\right], \\ H_{k\ell'}(y'|y_3) &:= \Pr\left[Y_{i4} \leq y' | \omega(i) = k, Y_{i3} = y_3, f(j_{i3}) = \ell', m_i = 1\right]. \end{aligned}$$

Moreover, in the first two periods we have:

$$\Pr\left[Y_{i1} \leq y | Y_{i2} = y_2, f(j_{i2}) = \ell\right] = \sum_{k=1}^{K} \pi_k\left(y_2, \ell\right) G_{k\ell}(y|y_2), \tag{25}$$

where:

$$\pi_k(y_2, \ell) := \Pr\left[\omega(i) = k | Y_{i2} = y_2, f(j_{i2}) = \ell\right].$$

Identification. Using (24), the extension of our setup which allows for non-stationary wage distributions directly provides identification of $G_{k\ell}(\cdot|y_2)$, $H_{k\ell'}(\cdot|y_3)$, and $p_k(y_2, y_3, \ell, \ell')$, almost surely in (y_2, y_3) and up to a common relabeling. To see this, note that (24) is formally analogous to (23), with y_2, ℓ (respectively y_3, ℓ') in place of ℓ (resp. ℓ'). Hence identification holds subject to the existence of suitable alternating cycles and rank conditions, as stated in Assumption 7. Note that in the present case the conditions are more demanding, as in particular they require all worker types to have positive probability, conditionally on both firm classes and wages.

Then, provided the $M \times K$ matrix with generic (y, k)-element $G_{k\ell}(y|y_2)$ has rank K almost surely in y_2, $\pi_k(y_2, \ell)$ is identified from (25) up to relabelling. Hence:

$$\pi_k(\ell) := \Pr[\omega(i) = k | f(j_{i2}) = \ell] = \mathbb{E}[\pi_k(Y_{i2}, \ell) | f(j_{i2}) = \ell]$$

is identified up to relabelling. By Bayes' rule, the period-2 cdf:

$$F_{k\ell}(y) := \Pr[Y_{i2} \leq y | \omega(i) = k, f(j_{i2}) = \ell] = \mathbb{E}\left[\frac{\pi_k(Y_{i2}, \ell)}{\pi_k(\ell)} \mathbf{1}\{Y_{i2} \leq y\} \middle| f(j_{i2}) = \ell\right]$$

is thus also identified up to relabelling. Similarly, the log wage cdfs in all other periods can be uniquely recovered up to relabelling, the period-3 and period-4 ones by making use of the bivariate distribution of (Y_{i3}, Y_{i4}). Transition probabilities associated with job change are also identified, as:

$$\Pr[f(j_{i3}) = \ell' | \omega(i) = k, Y_{i2} = y_2, f(j_{i2}) = \ell, m_i = 1] = \frac{\int p_k(y_2, y_3, \ell, \ell') q_{\ell\ell'}(y_2, y_3) dy_3}{\sum_{\widetilde{\ell}=1}^{L} \int p_k\left(y_2, y_3, \ell, \widetilde{\ell}\right) q_{\ell\widetilde{\ell}}(y_2, y_3) dy_3},$$

where $\int_{-\infty}^{y_3} q_{\ell\ell'}(y_2, y') dy' := \Pr[Y_{i3} \leq y_3, f(j_{i3}) = \ell' | Y_{i2} = y_2, f(j_{i2}) = \ell, m_i = 1]$.

Lastly, identification of the firm classes $f(j)$ follows as in the model with exogenous mobility and serial independence.

Estimation. A similar estimation approach to the one we described in Section 4 can be used here. In practice it is convenient to impose some structure on $p_k(y_2, y_3, \ell, \ell')$ in order to alleviate the curse of dimensionality. Once the firm classes have been estimated, we can maximize the following log-likelihood function:

$$\sum_{i=1}^{\widetilde{N}} \sum_{\ell=1}^{L} \sum_{\ell'=1}^{L} \mathbf{1}\{\widehat{f}(j_{i2}) = \ell\} \mathbf{1}\{\widehat{f}(j_{i3}) = \ell'\} \times \ln\left(\sum_{k=1}^{K} p_k(\ell, \ell') \varphi_{k\ell\ell'}(Y_{i2}, Y_{i3}; \eta) g_{k\ell}(Y_{i1}|Y_{i2}; \theta_1) h_{k\ell'}(Y_{i4}|Y_{i3}; \theta_2)\right),$$

with respect to θ_1, θ_2, η, and the $p_k(\ell, \ell')$, where the joint density $\varphi_{k\ell\ell'}(\cdot, \cdot; \eta)$ of (Y_{i2}, Y_{i3}) given $(\omega(i) = k, f(j_{i2}) = \ell, f(j_{i3}) = \ell', m_i = 1)$ is indexed by the parameter η. A practical approach is to specify a bivariate normal distribution with (k, ℓ, ℓ')-specific means and covariance matrix. To see the structure implicitly imposed on type probabilities, note that:

$$p_k(y_2, y_3, \ell, \ell'; \eta) = \frac{p_k(\ell, \ell') \varphi_{k\ell\ell'}(y_2, y_3; \eta)}{\sum_{k'=1}^{K} p_{k'}(\ell, \ell') \varphi_{k'\ell\ell'}(y_2, y_3; \eta)}.$$

Lastly, assuming a similar structure on:

$$\pi_k(y_2, \ell; \mu) = \frac{\pi_k(\ell) f_{k\ell}(y_2; \mu)}{\sum_{k'=1}^{K} \pi_{k'}(\ell) f_{k'\ell}(y_2; \mu)},$$

where we have parameterized the period-2 log wage density $f_{k\ell}(\cdot; \mu)$, the third estimation step consists in maximizing:

$$\sum_{i=1}^{N} \sum_{\ell=1}^{L} \mathbf{1}\{\widehat{f}(j_{i2}) = \ell\} \ln \left(\sum_{k=1}^{K} \pi_k(\ell) g_{k\ell}(Y_{i1} | Y_{i2}; \widehat{\theta}_1) f_{k\ell}(Y_{i2}; \mu) \right),$$

with respect to μ and the $\pi_k(\ell)$. In practice one can specify $f_{k\ell}(\cdot; \mu)$ as a normal distribution with (k, ℓ)-specific mean and variance.

The EM algorithm can be used for computation in the second and third estimation steps.

7.3 Identifying and estimating the model on $T > 2$ periods

To extend the model to more than 2 periods we need to model transitions between employment and unemployment, and the wage evolution of job stayers. Here we consider the version of the model that we described in Section 2, without endogenous mobility or serial correlation. As an example, consider job stayers between period 1 and 2 who move to another firm in period 3. Letting $m_{i1} = \mathbf{1}\{j_{i1} \neq j_{i2}\}$ and $m_{i2} = \mathbf{1}\{j_{i2} \neq j_{i3}\}$, we model:

$$\begin{aligned} &\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y', Y_{i3} \leq y'' \,|\, f(j_{i1}) = f(j_{i2}) = \ell, f(j_{i3}) = \ell', m_{i1} = 0, m_{i2} = 1\right] \\ &\qquad = \sum_{k=1}^{K} q_k(\ell, \ell') G_{k\ell}(y, y') F_{k\ell'}(y''), \end{aligned} \tag{26}$$

where $q_k(\ell, \ell')$ is the proportion of type-k workers who work in a class-ℓ firm in periods 1 and 2 and move to a class-ℓ' firm in period 3, and $G_{k\ell}(y, y')$ is the bivariate wage distribution of job stayers of type k in a class-ℓ firm. Transitions into and out of unemployment can be modelled in a similar way, under the assumption that wages are independent of past and future unemployment status conditional on worker type and firm classes, in addition to exogenous job

mobility. For example, consider job stayers between period 1 and 2 who lose their job in period 3. Let u_{i3} denote the unemployment indicator in period 3. We model:

$$\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = f(j_{i2}) = \ell, m_{i1} = 0, u_{i3} = 1\right] = \sum_{k=1}^{K} q_k(\ell, 0) G_{k\ell}(y, y'), \tag{27}$$

where $q_k(\ell, 0)$ is the proportion of type-k workers who work in a class-ℓ firm in periods 1 and 2 and lose their job in period 3.

The main difference with the two-period model concerns the identification of the bivariate distributions $G_{k\ell}$. Using similar notations as the proof of Theorem 1, (26) implies, for given y':

$$A(\ell, \ell', y') = G(\ell, y') Q(\ell, \ell') F(\ell')^{\intercal}.$$

As $F(\ell')$ is identified up to relabelling from Theorem 1, if follows that, for some permutation σ of $\{1, ..., K\}$, $q_{\sigma(k)}(\ell, \ell') G_{\sigma(k),\ell}(y, y')$ is identified, from which it follows that $G_{\sigma(k),\ell}$ is identified, provided $q_{\sigma(k)}(\ell, \ell') \neq 0$ for some ℓ'.

Lastly, for estimation, clustering and finite mixture estimators can be used similarly as in the main analysis.

7.4 Modelling of heterogeneity

Time-varying firm classes. Our framework can accommodate changes of firms classes over time, i.e., $f^t(j)$ instead of $f(j)$. This could reflect changes in firm productivity over time, or the impact of peer effects due to changes in worker composition (although note that distinguishing between different mechanisms would require additional assumptions). The only difference with the main analysis concerns the clustering stage when using job movers in both periods to infer firm classes; see (13). The finite mixture problem is identical to the one described in the main analysis.

Nonparametric clustering. Throughout we have fixed a finite number of firm classes L, and derived asymptotic results in this environment. An alternative approach would be to consider a setting where L grows at the same time as the number of workers and firms N and J. Several papers have recently considered this type of asymptotic framework in the statistical analysis of network data (Wolfe and Olhede, 2013; Gao, Lu, and Zhou, 2014).

Another possibility would be to adopt a local nonparametric approach, in the spirit of nearest neighbor matching. For any job transition of individual i from j_{i1} to j_{i2} one could construct the following neighborhood:

$$\left\{i', \text{ such that } j_{i',1} = j, j_{i',2} = j', \text{ and } \max\left(\left\|\widehat{F}_j - \widehat{F}_{j_{i1}}\right\|, \left\|\widehat{F}_{j'} - \widehat{F}_{j_{i2}}\right\|\right) \leq \delta\right\},$$

where $\widehat{F}_j$ denotes the empirical cdf of log wages in firm j in the cross-section, and $\delta > 0$ is a bandwidth parameter. The clustering approach that we use in this paper may similarly be understood as forming suitable neighborhoods, which allow to have sufficiently many job movements in order to infer worker and firm heterogeneity.

Continuous worker types. An alternative to our discrete modelling of worker heterogeneity is to assume that workers differ in a continuous type α_i, drawn from a class-conditional distribution $f(\alpha \,|\, \ell, \ell')$ given a job move between ℓ and ℓ'. One example is given by model (9). Here we consider a more general framework where wages are drawn from an unrestricted distribution $F(y|\alpha, \ell)$. Equations (5) and (6) then become:

$$\Pr\left[Y_{i1} \leq y\,,\, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell, f(j_{i2}) = \ell', m_i = 1\right] \;=\; \int f(\alpha \,|\, \ell, \ell') F(y|\alpha, \ell) F(y'|\alpha, \ell') d\alpha, \tag{28}$$

and:

$$\Pr\left[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell\right] = \int g\left(\alpha|\ell\right) F(y|\alpha, \ell) d\alpha. \tag{29}$$

The parameters of interest are the conditional wage distributions $F(y|\alpha, \ell)$ and the worker type composition $f(\alpha \,|\, \ell, \ell')$ and $g\left(\alpha|\ell\right)$. Nonparametric identification of these quantities based on (28) and (29) can be established using assumptions related to those in Hu and Schennach (2008), by relying on similar arguments as in Section 3. Estimation can then be based on parametric or nonparametric techniques, for example using the EM algorithm under parametric specifications. The identification and estimation of firm classes is identical as in the case with discrete worker heterogeneity.

8 Conclusion

In this paper we propose a framework for matched data. The clustering approach is useful in this context, as it allows to get back to standard single-agent econometric models. The idea to recover heterogeneous classes in a first step by clustering empirical distributions could be applied in other contexts, for example in structural models where joint estimation of the heterogeneity and the structural parameters might be computationally prohibitive.

Beyond the preliminary empirical results presented here, it would be interesting to document how worker and firm sorting varies with the business cycle and contributes to the evolution of wage inequality. Using our framework to model the earnings dynamics of workers across different

firms is another interesting agenda. Yet another possible application would be to estimate firm-level production functions, accounting for the worker composition heterogeneity that we have identified from wage data. Lastly, there are a number of questions to which matched data can provide interesting answers, and our method should be useful there as well. One example could be to use it to document sorting of teachers across schools.

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Discussion Paper 2002-06.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

Abowd, J. M., F. Kramarz, S. Pérez-Duarte, and I. Schmutte (2009): "A formal test of assortative matching in the labor market," Discussion paper, National Bureau of Economic Research.

Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008): "Mixed Membership Stochastic Blockmodels," *J. Mach. Learn. Res.*, 9, 1981–2014.

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?," *J. R. Stat. Soc. Ser. A Stat. Soc.*, 171(3), 673–697.

——— (2012): "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Econ. Lett.*, 117(3), 824–827.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J. M. Robin (2011): "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *Unpublished Manuscript*.

Bagger, J., and R. Lentz (2014): "An empirical model of wage dispersion with sorting," *University of Wisconsin mimeo*.

Becker, G. S. (1974): "A theory of marriage," *of the family: Marriage, children, and human capital*, pp. 299–351.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, 3, 993–1022.

Bonhomme, S., K. Jochmans, and J. M. Robin (2014): "Nonparametric estimation of finite mixtures from repeated measurements," *Journal of the Royal Statistical Society, Series B, forthcoming*.

Bonhomme, S., and E. Manresa (2012): "Grouped patterns of heterogeneity in panel data," Discussion paper, Citeseer.

Burdett, K., and D. T. Mortensen (1998): "Wage differentials, employer size, and unemployment," *Int. Econ. Rev.*, pp. 257–273.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality*," *Q. J. Econ.*, 128(3), 967–1015.

De Melo, R. L. (2009): "Sorting in the labor market: Theory and measurement," *Working Paper*.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. R. Stat. Soc. Series B Stat. Methodol.*, 39(1), 1–38.

Eeckhout, J., and P. Kircher (2011): "Identifying sorting in theory," *Rev. Econ. Stud.*, 78(3), 872–906.

Flinn, C., and J. Mabli (2008): "On-the-Job Search, Minimum Wages, and Labor Market Outcomes in an Equilibrium Bargaining," .

Friedrich, B., L. Laun, C. Meghir, and L. Pistaferri (2014): "Earnings Dynamics and Firm-Level Shocks," .

Gao, C., Y. Lu, and H. H. Zhou (2014): "Rate-optimal Graphon Estimation," .

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

Hagedorn, M., T. H. Law, and I. Manovskii (2014): "Identifying Equilibrium Models of Labor Market Sorting," *Working Paper*.

Hahn, J., and H. R. Moon (2010): "PANEL DATA MODELS WITH FINITE NUMBER OF MULTIPLE EQUILIBRIA," *Econometric Theory*, 26(03), 863–881.

Hall, P., and X.-H. Zhou (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *Ann. Stat.*, 31(1), 201–224.

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

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

Jackson, C. K. (2013): "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers," *Rev. Econ. Stat.*, 95(4), 1096–1116.

Kramarz, F., S. J. Machin, and A. Ouazad (2008): "What Makes a Test Score? The Respective Contributions of Pupils, Schools, and Peers in Achievement in English Primary Education," .

Lamadon, T., J. Lise, C. Meghir, and J. M. Robin (2013): "Matching, sorting and wages," *Working Paper*.

Levine, M., D. R. Hunter, and D. Chauveau (2011): "Maximum smoothed likelihood for multivariate mixtures," *Biometrika*.

Lin, C.-C., and S. Ng (2012): "Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown," *Journal of Econometric Methods*, 1(1).

Lise, J., C. Meghir, and J. M. Robin (2008): "Matching, sorting and wages," *Unpublished Manuscript*.

Mendes, R., G. J. van den Berg, and M. Lindeboom (2010): "An empirical assessment of assortative matching in the labor market," *Labour Econ.*, 17(6), 919–929.

Moscarini, G. (2005): "Job Matching and the Wage Distribution," *Econometrica*, 73(2), 481–516.

Reiersøl, O. (1950): "Identifiability of a Linear Relation between Variables Which Are Subject to Error," *Econometrica*, 18(4), 375–389.

Shimer, R. (2006): "On-the-job search and strategic bargaining," *Eur. Econ. Rev.*, 50(4), 811–830.

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

Steinley, D. (2006): "K-means clustering: a half-century synthesis," *Br. J. Math. Stat. Psychol.*, 59(1), 1–34.

Wolfe, P. J., and S. C. Olhede (2013): "Nonparametric graphon estimation," .

Woodcock, S. D. (2008): "Wage differentials in the presence of unobserved worker, firm, and match heterogeneity," *Labour Econ.*, 15(4), 771–793.

APPENDIX

A Proofs

Proof of Theorem 1. Let $\ell \in \{1, ..., L\}$, and let $(\ell_1, ..., \ell_{2R})$ from Assumption 4. From (5) we have, considering workers who move between ℓ_r and ℓ_{r+1} for some $r \in \{1, ..., R\}$:

$$\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell_r, f(j_{i2}) = \ell_{r+1}, m_i = 1\right] = \sum_{k=1}^{K} p_k(\ell_r, \ell_{r+1}) F_{k,\ell_r}(y) F_{k,\ell_{r+1}}(y'). \quad (30)$$

Consider a set of M values for y, and the same set of values for y', that satisfy Assumption 4 *ii*). Writing (30) in matrix notation we obtain

$$A(\ell_r, \ell_{r+1}) = F(\ell_r) D(\ell_r, \ell_{r+1}) F(\ell_{r+1})^\intercal, \quad (31)$$

where $A(\ell_r, \ell_{r+1})$ is $M \times M$ with generic element:

$$\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell_r, f(j_{i2}) = \ell_{r+1}, m_i = 1\right],$$

$F(\ell_r)$ is $M \times K$ with element $F_{k,\ell_r}(y)$, $F(\ell_{r+1})$ is $M \times K$ with element $F_{k,\ell_{r+1}}(y')$, and $D(\ell_r, \ell_{r+1})$ is $K \times K$ diagonal with element $p_k(\ell_r, \ell_{r+1})$.

Note that $A(\ell_r, \ell_{r+1})$ has rank K by Assumption 4 *ii*). Consider a singular value decomposition of $A(\ell_1, \ell_2)$:

$$F(\ell_1) D(\ell_1, \ell_2) F(\ell_2)^\intercal = USV^\intercal,$$

where S is $K \times K$ diagonal and non-singular, and U and V have orthonormal columns.

We define the following matrices:

$$\begin{aligned} B(\ell_r, \ell_{r+1}) &:= S^{-\frac{1}{2}} U^\intercal A(\ell_r, \ell_{r+1}) V^\intercal S^{-\frac{1}{2}}, \\ C(\ell_r, \ell_{r+1}) &:= S^{-\frac{1}{2}} U^\intercal \left[A(\ell_r, \ell_{r+1})^\intercal\right] V^\intercal S^{-\frac{1}{2}}, \\ Q(\ell_r) &:= S^{-\frac{1}{2}} U^\intercal F(\ell_r). \end{aligned}$$

Note that $B(\ell_r, \ell_{r+1})$, $C(\ell_r, \ell_{r+1})$, and $Q(\ell_r)$ are non-singular, by Assumption 4 *ii*). Moreover, we have, for all $r \in \{1, ..., 2R-1\}$:

$$\begin{aligned} B(\ell_r, \ell_{r+1}) C(\ell_{r+1}, \ell_{r+2})^{-1} &= S^{-\frac{1}{2}} U^\intercal A(\ell_r, \ell_{r+1}) V^\intercal S^{-\frac{1}{2}} \left(S^{-\frac{1}{2}} U^\intercal A(\ell_{r+1}, \ell_{r+2})^\intercal V^\intercal S^{-\frac{1}{2}}\right)^{-1} \\ &= S^{-\frac{1}{2}} U^\intercal F(\ell_r) D(\ell_r, \ell_{r+1}) \left(S^{-\frac{1}{2}} U^\intercal F(\ell_{r+2}) D(\ell_{r+1}, \ell_{r+2})\right)^{-1} \\ &= Q(\ell_r) D(\ell_r, \ell_{r+1}) D(\ell_{r+1}, \ell_{r+2})^{-1} Q(\ell_{r+2})^{-1}. \end{aligned}$$

Let $E_r = B(\ell_r, \ell_{r+1}) C(\ell_{r+1}, \ell_{r+2})^{-1}$. We thus have:

$$E_1 E_3 ... E_{2R-1} = Q(\ell_1) D(\ell_1, \ell_2) D(\ell_2, \ell_3)^{-1} ... D(\ell_{2R-1}, \ell_{2R}) D(\ell_{2R}, \ell_1)^{-1} Q(\ell_1)^{-1}.$$

The eigenvalues of this matrix are all distinct by Assumption 4 i), so $Q(\ell_1)$ is identified up to multiplication by a diagonal matrix and permutation of its columns. As $\ell_1 = \ell$, this shows that $Q(\ell)$ is identified up to column scaling and permutation.

Now, note that $F(\ell) = UU^\intercal F(\ell)$, so:

$$F(\ell) = US^{\frac{1}{2}}Q(\ell)$$

is identified up to multiplication by a diagonal matrix and permutation of its columns. Hence $F_{k\ell}(y)\lambda_k$ is identified up to relabelling, where $\lambda_k \neq 0$ is a scale factor. As a result, $\lambda_k = \lim_{y\to+\infty} F_{k\ell}(y)\lambda_k$ is identified, so $F_{k\ell}$ is identified up to relabelling.

Let now $\ell' \neq \ell$, and let $(\ell_1, ..., \ell_R)$ be a cycle containing $\ell = \ell_1$ and $\ell' = \ell_{\overline{r}}$, obtained from Assumption 4. We have:

$$A(\ell, \ell_2) = F(\ell)D(\ell, \ell_2)F(\ell_2)^\intercal. \tag{32}$$

The above arguments show that there exists a permutation $\sigma : \{1, ..., K\} \to \{1, ..., K\}$ such that $F_{\sigma(k),\ell}$ is identified. Hence, by (32):

$$p_{\sigma(k)}(\ell, \ell_2)F_{\sigma(k),\ell_2}(y')$$

is identified, so by taking limits as $y' \to +\infty$ both $p_{\sigma(k)}(\ell, \ell_2)$ and $F_{\sigma(k),\ell_2}$ are identified. By induction, $p_{\sigma(k)}(\ell_r, \ell_{r+1})$ and $F_{\sigma(k),\ell_{r+1}}$ are identified for all r. As $\ell' = \ell_{\overline{r}}$, it follows that $F_{\sigma(k),\ell'}$ is identified.

Lastly, let any $(\ell, \ell') \in \{1, ..., L\}^2$. Then, from:

$$A(\ell, \ell') = F(\ell)D(\ell, \ell')F(\ell')^\intercal,$$

$p_{\sigma(k)}(\ell, \ell')$ is identified as $F_{\sigma(k),\ell}$ and $F_{\sigma(k),\ell'}$ are both identified, and as $F(\ell)$ and $F(\ell')$ have rank K by Assumption 4 ii).

Proof of Corollary 1. By Assumptions 1, 2, 3, 4, and Theorem 1 there exists a permutation $\sigma : \{1, ..., K\} \to \{1, ..., K\}$ such that $F_{\sigma(k),\ell}$ is identified for all ℓ. Now we have, writing (6) for the M values of y given by Assumption 4 ii) in matrix form:

$$H(\ell) = F(\ell)\Pi(\ell),$$

where $H(\ell)$ has generic element $\Pr[Y_{i1} \leq y \,|\, f(j_{i1}) = \ell]$, the $K \times 1$ vector $\Pi(\ell)$ has generic element $\pi_{\sigma(k)}(\ell)$, and the columns of $F(\ell)$ have been ordered with respect to σ. By Assumption 4 ii) $F(\ell)$ has rank K, from which it follows that:

$$\Pi(\ell) = \left[F(\ell)^\intercal F(\ell)\right]^{-1} F(\ell)^\intercal H(\ell)$$

is identified. So $\pi_{\sigma(k)}(\ell)$ is identified.

Proof of Proposition 1. Let $j \in \{1, ..., J\}$ such that $f(j) = \ell$. For all firms $j'' \neq j$ such that $f(j'') = \ell''$, the bivariate log wage distribution of job movers from j to j'', say $F_{jj''}$, is equal to $G_{\ell\ell''}$, which is given by (13). Similarly, the bivariate log wage distribution of job movers from j'' to j, $F_{j''j}$, is equal to $G_{\ell''\ell}$. Hence, by Assumption 5, the set:

$$\mathcal{S}_j := \left\{ \left(F_{jj''}, F_{j''j} \right), j'' \neq j \right\}$$

takes exactly L values. The $f(j)$ are thus identified as the equivalence classes of $\mathcal{S}_j$.

Note that the number of classes L is also identified, as the number of points of support of $\mathcal{S}_j$.

Proof of Corollary 2. Using similar notations as in the proof of Theorem 1, we have:

$$F(\ell)D(\ell, \ell'')F(\ell'')^\intercal = F(\ell')D(\ell', \ell'')F(\ell'')^\intercal.$$

By Assumption 4 *ii*), $F(\ell'')$ has rank K. So:

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

Using similar arguments as in the proof of Theorem 1 we thus obtain that $F_{k\ell} = F_{k\ell'}$ for all k.

Proof of Proposition 2. We verify Assumptions 1 and 2 in Bonhomme and Manresa (2012). Note that here there are no covariates. Assumptions 1a and 1c are satisfied because $\mathbf{1}\{Y_{i1} \leq y\}$ is bounded. Assumptions 1d, 1e, and 1f hold because of Assumption 6 *i*). Assumptions 2a and 2b hold by Assumptions 6 *ii*) and *iii*). Finally, Assumptions 2c and 2d are also satisfied by Assumption 6 *i*) and boundedness of $\mathbf{1}\{Y_{i1} \leq y\}$. Theorems 1 and 2 in Bonhomme and Manresa (2012) and Assumption 6 *iv*) imply the result.

Proof of Theorem 2. Let $\ell \in \{1, ..., L\}$, and let $(\ell_1, ..., \ell_R)$, $(\widetilde{\ell}_1, ..., \widetilde{\ell}_R)$ as in Assumption 7. In particular, $\ell_1 = \ell$. As in the proof of Theorem 1 we have that, for all r and all $r' \in \{r-1, r\}$:

$$A(\ell_r, \widetilde{\ell}_{r'}) = F^1(\ell_r) D^{12}(\ell_r, \widetilde{\ell}_{r'}) F^2(\widetilde{\ell}_{r'})^\intercal,$$

have rank K, where $A(\ell_r, \widetilde{\ell}_{r'})$ is $M \times M$ with generic element:

$$\Pr\left[Y_{i1} \leq y, Y_{i2} \leq y' \,|\, f(j_{i1}) = \ell_r, f(j_{i2}) = \widetilde{\ell}_{r'}, m_i = 1 \right],$$

$F^1(\ell_r)$ is $M \times K$ with element $F^1_{k,\ell_r}(y)$, $F^2(\widetilde{\ell}_{r'})$ is $M \times K$ with element $F^2_{k,\widetilde{\ell}_{r'}}(y')$, and $D^{12}(\ell_r, \widetilde{\ell}_{r'})$ is $K \times K$ diagonal with element $p^{12}_k(\ell_r, \widetilde{\ell}_{r'})$.

Consider a singular value decomposition of $A(\ell_1, \widetilde{\ell}_1)$:

$$F^1(\ell_1) D^{12}(\ell_1, \widetilde{\ell}_1) F^2(\widetilde{\ell}_1)^\intercal = USV^\intercal,$$

where S is $K \times K$ diagonal and non-singular, and U and V have orthonormal columns. We define the following matrices:

$$\begin{aligned} B(\ell_r, \widetilde{\ell}_{r'}) &= S^{-\frac{1}{2}} U^\intercal A(\ell_r, \widetilde{\ell}_{r'}) V^\intercal S^{-\frac{1}{2}}, \\ Q(\ell_r) &= S^{-\frac{1}{2}} U^\intercal F^1(\ell_r). \end{aligned}$$

$B(\ell_r, \widetilde{\ell}_{r'})$ and $Q(\ell_r)$ are non-singular by Assumption 7 *ii*). Moreover, we have, for all $r \in \{1, ..., R\}$:

$$\begin{aligned} B(\ell_r, \widetilde{\ell}_r) B(\ell_{r+1}, \widetilde{\ell}_r)^{-1} &= S^{-\frac{1}{2}} U^\intercal A(\ell_r, \widetilde{\ell}_r) V^\intercal S^{-\frac{1}{2}} \left(S^{-\frac{1}{2}} U^\intercal A(\ell_{r+1}, \widetilde{\ell}_r) V^\intercal S^{-\frac{1}{2}} \right)^{-1} \\ &= S^{-\frac{1}{2}} U^\intercal F^1(\ell_r) D^{12}(\ell_r, \widetilde{\ell}_r) \left(S^{-\frac{1}{2}} U^\intercal F^1(\ell_{r+1}) D^{12}(\ell_{r+1}, \widetilde{\ell}_r) \right)^{-1} \\ &= Q(\ell_r) D^{12}(\ell_r, \widetilde{\ell}_r) D^{12}(\ell_{r+1}, \widetilde{\ell}_r)^{-1} Q(\ell_{r+1})^{-1}. \end{aligned}$$

Let $E_r = B(\ell_r, \widetilde{\ell}_r) B(\ell_{r+1}, \widetilde{\ell}_r)^{-1}$. We thus have:

$$E_1 E_2 ... E_R = Q(\ell_1) D^{12}(\ell_1, \widetilde{\ell}_1) D^{12}(\ell_2, \widetilde{\ell}_1)^{-1} ... D^{12}(\ell_R, \widetilde{\ell}_R) D^{12}(\ell_1, \widetilde{\ell}_R)^{-1} Q(\ell_1)^{-1}.$$

The eigenvalues of this matrix are all distinct by Assumption 7 *i*), so $Q(\ell_1) = Q(\ell)$ is identified up to multiplication by a diagonal matrix and permutation of its columns. It follows as in the proof of Theorem 1 that $F^1_{\sigma(k),\ell}$ is identified for some permutation $\sigma : \{1, ..., K\} \to \{1, ..., K\}$.

Let now $\ell' \neq \ell$, and let $(\ell_1, ..., \ell_R)$, $(\widetilde{\ell}_1, ..., \widetilde{\ell}_R)$, be an alternating cycle such that $\ell_1 = \ell$ and $\ell' = \ell_r$ for some r, by Assumption 7 *i*). It follows as in the proof of Theorem 1 that $F^1_{\sigma(k),\ell'}$ is identified.

Next, let ℓ, ℓ', possibly equal, and let $(\ell'_1, ..., \ell'_R)$, $(\widetilde{\ell}'_1, ..., \widetilde{\ell}'_R)$ be an alternating cycle such that $\ell'_1 = \ell$ and $\ell' = \widetilde{\ell}'_r$ for some r, by Assumption 7 *i*). We have:

$$B(\ell'_1, \widetilde{\ell}'_1) B(\ell'_2, \widetilde{\ell}'_1)^{-1} ... B(\ell'_r, \widetilde{\ell}'_r) = Q(\ell'_1) D^{12}(\ell'_1, \widetilde{\ell}'_1) D^{12}(\ell'_2, \widetilde{\ell}'_1)^{-1} ... D^{12}(\ell'_r, \widetilde{\ell}'_r) F^2(\widetilde{\ell}'_r)^\intercal V^\intercal S^{-\frac{1}{2}},$$

from which it follows, similarly as in the proof of Theorem 1, that $F^2_{\sigma(k),\ell'} = F^2_{\sigma(k),\widetilde{\ell}'_r}$ is identified.

Lastly, let any $(\ell, \ell') \in \{1, ..., L\}^2$. Then, from:

$$A(\ell, \ell') = F^1(\ell) D^{12}(\ell, \ell') F^2(\ell')^\intercal,$$

$p^{12}_{\sigma(k)}(\ell, \ell')$ is identified as $F^1_{\sigma(k),\ell}$ and $F^2_{\sigma(k),\ell'}$ are both identified, and as $F^1(\ell)$ and $F^2(\ell')$ have rank K by Assumption 7 *ii*).

B Links to a theoretical model

In this section of the appendix we consider a variation of the model of Shimer and Smith (2000) with on-the-job search. Relative to the main text we modify some of the notation, in order to be closer to the original paper.

Environment. The economy is composed of a uniform measure of workers indexed by x with unit mass and a uniform measure of jobs indexed by y with mass $\bar{V}$. A match (x, y) produces output $f(x, y)$ and separates exogenously at rate δ. Workers are employed or unemployed. We denote $u(x)$ the measure of unemployed, $h(x, y)$ the measure of matches, and $v(y)$ the measure of vacancies. We let $U = \int u(x)\mathrm{d}x$ the mass of unemployed, and $V = \int v(y)\mathrm{d}y$ the mass of vacancies. Unemployed workers meet vacancies at rate λ_0, and employed workers meet vacancies at rate λ_1. Vacancies meet unemployed workers at rate μ_0, and employed workers at rate μ_1. A firm cannot advertise for a job that is currently filled. Unemployed workers collect benefits $b(x)$, and vacancies have to pay a flow cost $c(y)$.

Timing. Each period is divided into four stages. In stage one, active matches collect output and pay wages. In stage two, active matches exogenously separate with probability δ. In stage three vacant jobs can advertise and all workers can search. In stage four workers and vacant jobs meet randomly and, upon meeting, the worker and the firm must decide whether or not to match based on expected surplus generated by the match. The wage is set by Nash bargaining, where α is the bargaining power of the worker. We assume that wages are continuously renegotiated with the value of unemployment (see Shimer (2006) for a discussion). Since workers and firms can search in the same period as job losses occur, it is convenient to introduce within periods distributions:

$$v_{1/2}(y) := \frac{\delta + (1-\delta)v(y)}{\delta + (1-\delta)V}, \quad u_{1/2}(x) := \frac{\delta + (1-\delta)u(x)}{\delta + (1-\delta)U}, \quad h_{1/2}(x, y) := \frac{h(x, y)}{1-U},$$

where each distribution integrates to one by construction.

Value functions. We then write down the value functions for this model. Let $S(x, y)$ be the surplus of the match, $W_0(x)$ the value of unemployment, and $\Pi_0(y)$ the value of a vacancy. We have:

$$rW_0(x) = (1+r)b(x) + \lambda_0 \int M(x, y)\alpha S(x, y)v_{1/2}(y)\mathrm{d}y, \tag{BE-W0}$$

and:

$$r\Pi_0(y) = \mu_0 \int M(x, y)(1-\alpha)S(x, y)u_{1/2}(x)\mathrm{d}x + \mu_1 \iint P(x, y', y)(1-\alpha)S(x, y)h_{1/2}(x, y')\mathrm{d}y'\mathrm{d}x, \tag{BE-P0}$$

where $M(x, y) := \mathbf{1}\{S(x, y) \geq 0\}$ is the matching decision, and $P(x, y', y)$ is one when $S(x, y) > S(x, y')$ (that is, when y is preferred to y' by x), zero when $S(x, y) < S(x, y')$, and $1/2$ when $S(x, y) = S(x, y')$.

We write the Bellman equation for a job y that currently employs a worker x at wage w:

$$\begin{aligned}(r+\delta)\Pi_1(x, y, w) = (1+r)\left[f(x, y) - w + \delta\left(\Pi_0(y) + c(y)\right)\right] \\ - (1-\delta)\lambda_1 q(x, y)(1-\alpha)S(x, y),\end{aligned}$$

where $q(x,y) = \int P(x,y,y')v_{1/2}(y')\mathrm{d}y'$ represents the total proportion of firms y' that can poach a worker x from firm y. We then turn to the Bellman equation for the employed worker:

$$(r+\delta)W_1(x,y,w) = (1+r)\left[w + \delta\left(W_0(x) - b(x)\right)\right] + (1-\delta)\lambda_1 \int P(x,y,y')(\alpha S(x,y') - \alpha S(x,y))v_{1/2}(y')\mathrm{d}y'. \tag{BE-W1}$$

Finally, we derive the value of the surplus associated with the match (x,y), defined by $S := W_1 + \Pi_1 - W_0 - \Pi_0$:

$$(r+\delta)S(x,y) = (1+r)\left[f(x,y) - \delta\left(b(x) - c(y)\right)\right] - r(1-\delta)\left(\Pi_0(y) + W_0(x)\right) + (1-\delta)\lambda_1 \int P(x,y,y')(\alpha S(x,y') - S(x,y))v_{1/2}(y')\mathrm{d}y'. \tag{BE-S}$$

Flow equations. Lastly we write the flow equation for the joint distribution of matches at the beginning of the period:

$$\big(\delta + (1-\delta)\lambda_1 q(x,y)\big)h(x,y) = \lambda_0\left(\delta + (1-\delta)U\right)u_{1/2}(x)v_{1/2}(y)M(x,y) + \lambda_1(1-\delta)(1-U)\int P(x,y',y)h_{1/2}(x,y')\mathrm{d}y' v_{1/2}(y), \tag{EQ-H}$$

where:

$$\mu_0\left(\delta + (1-\delta)V\right) = \lambda_0\left(\delta + (1-\delta)U\right), \text{ and } \mu_1\left(\delta + (1-\delta)V\right) = \lambda_1(1-\delta)(1-U), \tag{MC-S}$$

are the total number of matches coming out of unemployment and coming from on-the-job transitions, respectively. The market clearing conditions on the labor market are given by:

$$\int h(x,y)\mathrm{d}x + v(y) = \bar{V}, \text{ and } \int h(x,y)\mathrm{d}y + u(x) = 1. \tag{MC-L}$$

Equilibrium. For a set of primitives $\delta, \lambda_0, \lambda_1, f(x,y), b(x), c(y), \alpha$, the stationary equilibrium is characterized by the values $S(x,y), W_0(x), \Pi_0(y)$ and the measure of matches $h(x,y)$ such that i) Bellman equations (BE-W0), (BE-P0) and (BE-S) are satisfied, ii) h satisfies the flow equation (EQ-H), and iii) the constraints (MC-S) and (MC-L) hold.

Wage. We then derive the wage from the model using equation (BE-W1) and using that Nash bargaining gives $W_1(x,y,w(x,y)) = \alpha S(x,y) + W_0(x)$:

$$(1+r)w(x,y) = (r+\delta)\alpha S(x,y) + (1-\delta)rW_0(x) - (1-\delta)\lambda_1 \int P(x,y,y')(\alpha S(x,y') - \alpha S(x,y))v_{1/2}(y')\mathrm{d}y'.$$

Mapping to distributional model. From there we can recover our model's cross sectional worker type proportions conditional on firm heterogeneity ($\pi_k(\ell)$ in the body of the paper):

$$\pi(\omega = x|f = y) = \frac{h(x,y)}{1 - v(y)},$$

and the type proportions for job movers ($p_k(\ell, \ell')$ in the main text), which are given by:

$$p(\omega = x|f = y, f' = y') = \frac{\big(\delta\lambda_0 + (1-\delta)\lambda_1 \mathbf{1}\{S(x,y') > S(x,y)\}\big)h(x,y)M(x,y')}{\int \big(\delta\lambda_0 + (1-\delta)\lambda_1 \mathbf{1}\{S(\tilde{x},y') > S(\tilde{x},y)\}\big)h(\tilde{x},y)M(\tilde{x},y')\mathrm{d}\tilde{x}}.$$

Lastly, we assume that the wage is measured with a multiplicative independent measurement error:

$$\tilde{w} = w(x,y)\exp(\varepsilon),$$

from which we can derive the marginal log-wage distributions ($F_{k\ell}$ in the main text).

Without on-the-job search ($\lambda_1 = \mu_1 = 0$). Let us consider the case without on-the-job search. Equation (EQ-H) gives:

$$\delta h(x,y) = \lambda_0 \left(\delta + (1-\delta)U\right) u_{1/2}(x) v_{1/2}(y) M(x,y).$$

Hence:

$$p(\omega = x|f = y, f' = y') = \frac{M(x,y)M(x,y')u_{1/2}(x)}{\int M(\tilde{x},y)M(\tilde{x},y')u_{1/2}(\tilde{x})\mathrm{d}\tilde{x}}. \tag{PX-YY'}$$

These probabilities are symmetric in (y, y'). In the context of Theorem 1 this means that Assumption 4 *i*) is not satisfied, as $a_k = 1$ for all k. This is the setup considered in Shimer and Smith (2000) and Hagedorn, Law, and Manovskii (2014), for example. Symmetry occurs because, in these models, all job changes are associated with an intermediate unemployment spell, where all information about the previous firm disappears. Empirically the majority of job changes occur via job-to-job transitions. Moreover, in Figure 2 we find evidence against the particular symmetry of equation (PX-YY').

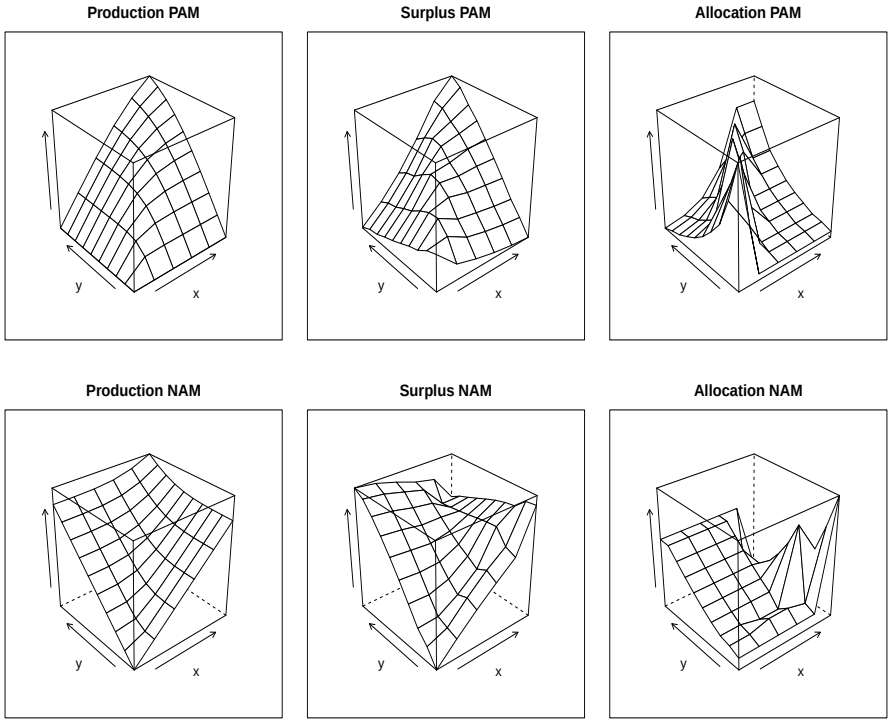
Simulation and estimation. We pick two parameterizations of the model associated with positive assortative matching (PAM) and negative assortative matching (NAM) in equilibrium. We set $b(x) = b = 0.3$, $c(y) = c = 0$, and $\bar{V} = 2$. We solve the model at a yearly frequency, and we set $\delta = 0.02$, $\lambda_0 = 0.4$ and $\lambda_1 = 0.1$. The production function is CES:

$$f(x,y) = a + \left(\nu x^\rho + (1-\nu)y^\rho\right)^{1/\rho},$$

where we set $\nu = 0.5$ and $a = 0.7$. Finally we consider $\rho = -3$ (PAM) and $\rho = 4$ (NAM).

Figures 13 and 14 plot the model solutions, in terms of production, surplus, allocation, and log wages. We see clear differences between PAM and NAM. In particular, in NAM mean log wages are not monotone in firm productivity.

Figure 13: Model solutions: production, surplus and allocation

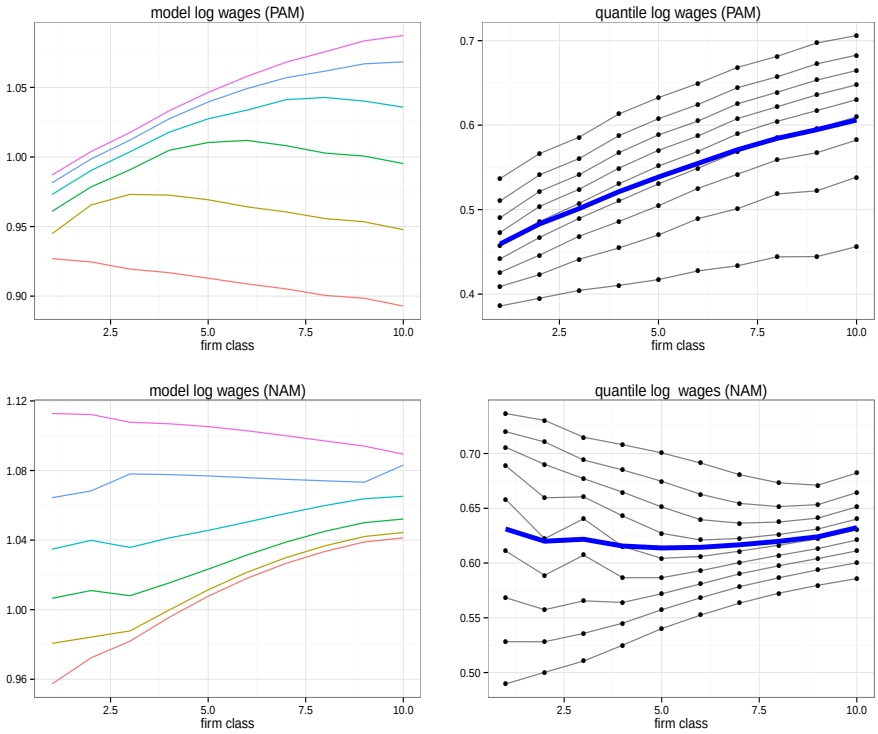


Notes: The graphs show the model solution in terms of production $f(x, y)$, surplus $S(x, y)$, and allocation $h(x, y)$. Positive assortative matching (top panel), and negative assortative matching (bottom panel).

In Table 4 we report the results of variance decompositions based on the data generated according to the model, and based on estimates from our distributional model based on those data. We use $K = 6$ and $L = 10$ in estimation, and consider two scenarios for K and L in the model: $(6, 10)$ and $(50, 50)$. In the first three columns we show the results of a decomposition of the variance of log wages in terms of between-worker, within-worker between-firm, and within-worker within-firm components. We see in the first four rows that, when we use as many heterogeneity types in estimation as in the true model our approach tends to underestimate the worker contribution and overestimate the within-worker-and-firm component. Nevertheless, the decomposition is rather well reproduced, especially taking into account that we re-estimate the firm classes together with all model parameters. In the last four rows of the table the results are comparable, although for NAM the within-worker between-firm contribution is underestimated. This suggests that our approach may still provide informative answers in situations where worker and firm heterogeneity are not clustered into a few classes and types.

In the last four columns of Table 4 we show the results of additive variance decomposition exercises, similar to the ones reported in Table 3. We see that our approach recovers the worker and firm

Figure 14: Model solutions: log wages



Notes: The left graphs show log wages (without measurement error), by worker type and firm class. The right graphs show deciles of log wages (with measurement error) by firm class. The thick lines correspond to mean log wages. Positive assortative matching (top panel), and negative assortative matching (bottom panel).

components rather well, even when the number of types used in estimation, $(6, 10)$, is smaller than the true one, $(50, 50)$. In addition, our approach correctly recovers the sign of the correlation between worker and firm effects, for both positive and negative assortative matching. Moreover, the magnitude of the correlation is very well estimated for PAM, and slightly less so for NAM, the bias being highest in the last row of the table.

Table 4: Variance decompositions on data generated by a theoretical model

		dim	%bw	%wwbf	%wwwf	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$
PAM	model	6×10	0.693	0.103	0.203	0.791	0.054	0.156	0.377
	BLM	6×10	0.636	0.101	0.263	0.756	0.069	0.175	0.385
NAM	model	6×10	0.661	0.136	0.203	1.082	0.125	-0.206	-0.281
	BLM	6×10	0.625	0.114	0.262	1.049	0.099	-0.148	-0.23
PAM	model	50×50	0.693	0.108	0.2	0.758	0.071	0.171	0.367
	BLM	6×10	0.591	0.121	0.288	0.701	0.095	0.204	0.396
NAM	model	50×50	0.685	0.115	0.201	1.079	0.107	-0.186	-0.273
	BLM	6×10	0.668	0.044	0.288	1.009	0.041	-0.05	-0.122

Notes: Variance decompositions based on data generated from the theoretical sorting model with PAM or NAM. "BLM" corresponds to estimates based on our approach. "dim" is $K \times L$, where K is the number of worker types and L is the number of firm classes. "%bw", "%wwbf", and "%wwwf" denote the between-worker, within-worker between-firm, and within-worker within-firm components of the variance of log wages, respectively. The last four columns correspond to an additive variance decomposition, similar to Table 3.