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: February 15, 2016

IN PROGRESS

Abstract

We propose a framework to estimate earnings distributions and worker and firm unobserved heterogeneity on matched panel data. We introduce two models: a static model that allows for interaction effects between workers and firms, and a dynamic model that allows in addition for Markovian dynamics in earnings and mobility decisions. We establish identification in short panels. We develop a tractable two-step estimation approach where firms are classified into heterogeneous classes in a first step. Finally, we apply our method to Swedish matched employer employee panel data and report estimated earnings functions, sorting patterns, and variance decompositions.

JEL codes: J31, J62, C23.

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

*We thank our discussants Pat Kline and Chris Hansen. We also thank Rafael Lopes de Melo, Costas Meghir, Luigi Pistaferri, Jean-Marc Robin, Adam Rosen, and Robert Shimer for useful comments. We thank the IFAU for access to, and help with, the Swedish administrative data. The usual disclaimer applies. Contact: sbonhomme@uchicago.edu; lamadon@uchicago.edu; emanresa@mit.edu .

1 Introduction

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

Two influential literatures have approached these questions from different angles. The method of Abowd, Kramarz, and Margolis (1999) (AKM hereafter) relies on two-way fixed-effect regressions to account for unobservable worker and firm effects, and allows quantifying their respective contributions to earnings dispersion, as well as the correlation between worker and firm unobservables. The AKM method has had a very large impact, in labor economics and outside.[1] A second literature has tackled similar issues from a structural perspective, by developing and estimating fully specified theoretical models of sorting on the labor market.[2]

Reconciling these literatures has proven difficult, however. While the AKM method provides a tractable way to deal with two-sided unobserved heterogeneity, the AKM model relies on substantive, possibly restrictive assumptions. The absence of interaction terms between worker and firm attributes restricts the patterns of complementarity in earnings. Since Becker's work, numerous theories have emphasized the link between complementarity and sorting (Shimer and Smith, 2000, Eeckhout and Kircher, 2011). Moreover, the AKM model is static in the sense that worker mobility may not depend on earnings directly conditional on worker's and firm's fixed unobservables, and that earnings after a job move are assumed not to depend on the previous firm's attributes. These static aspects may conflict with implications of dynamic economic models.[3]

On the other hand, attempts at structurally estimating dynamic models of sorting have faced computational and empirical challenges. The dimensions involved are daunting: how to estimate a model of workers' mobility and earnings with hundreds of thousands of workers and

[1] Applications of the method to earnings data include Gruetter and Lalive (2009), Mendes, van den Berg, and Lindeboom (2010), Woodcock (2008), Card, Heining, and Kline (2013), Goldschmidt and Schmieder (2015), and Song, Price, Guvenen, and Bloom (2015), 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, or to document differences across areas in patients' health care utilization (Kramarz, Machin, and Ouazad, 2008, Jackson, 2013, Finkelstein, Gentzkow, and Williams, 2014).

[2] Many structural models proposed in the literature build on Becker (1973). Examples are 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), and Bagger and Lentz (2014).

[3] For example, they may not be consistent with wage posting models with match-specific heterogeneity, or with sequential auctions mechanisms as in Postel-Vinay and Robin (2002), as we discuss below.

dozens of thousands of firms in the presence of both firm and worker unobserved heterogeneity? And how informative are functional forms assumptions in these often tightly parameterized models?

In this paper we take a step in the direction of building a bridge between these two approaches. We propose two empirical models, static and dynamic, which allow for several aspects that are absent in two-way fixed-effects regressions. In the static model we allow for interaction effects between worker and firm unobservables. In the dynamic model we let job mobility depend on earnings directly, in addition to worker and firm attributes. Moreover, we allow earnings after a job move to depend on attributes of the previous firm, in addition to those of the current one. The model exhibits first-order Markov dynamics, thus restricting the dimension of the relevant state space. While we do not take a structural route in this paper, we show that our assumptions nest several theoretical mechanisms that have been emphasized in the literature.

We provide formal conditions for identification, under discrete and continuous worker heterogeneity. For the static model we rely on two periods, while we need four periods to identify the dynamic model. The ability of the method to deal with short panels is important, since assuming time-invariant effects over long periods may be strong. This also opens the way to document how earnings premia and sorting patterns vary over the business cycle. In addition, although we focus on workers and firms in this paper, our framework could be useful in other applications using matched data, such as teacher-student sorting, where long panels may not be available. Our results emphasize that, in order to identify models with complementarity, one requires workers of different types to make heterogeneous transitions between firms. This generalizes the intuitive observation that worker/firm interaction effects would not be identified if workers' allocation to firms was fully random.

In the models, the relevant level of firm unobserved heterogeneity is the *class* of the firm. In principle, these classes could be the firms themselves, in applications with very large firms or in sorting applications to other settings (e.g., across cities). However, in matched employer employee panel data sets of typical sizes, implementing our approach requires reducing the number of classes. This dimension reduction serves two purposes: it helps computation and, importantly, it allows increasing statistical precision and alleviating the incidental parameter bias caused by the presence of a very large number of firm-specific parameters.[4] We use a clustering estimator to classify firms into classes based on how similar their earnings distributions are.

[4]See Andrews, Gill, Schank, and Upward (2008, 2012) for illustrations of incidental parameter bias in two-way fixed-effects regressions.

The clustering could also be based on mobility patterns or longitudinal earnings information, and it could be modified to incorporate firm characteristics such as size or value-added. We establish the asymptotic consistency of the classification under discrete firm heterogeneity by verifying the conditions of the main theorems in Bonhomme and Manresa (2015). Under these conditions, estimation error in classification does not affect inference on parameters estimated in a second step.[5] This provides a formal justification for clustering in our models.

We propose a two-step approach to estimate the models. In the first step we classify firms into classes using clustering. The second step depends on the model considered. As an example, we consider static and dynamic extensions of the AKM regression model that allow for interaction effects between firms and workers. The second step then takes the form of simple linear instrumental variables and covariance-based estimation, conditional on the firm classes. We also consider finite mixture models where worker types are assumed to be discrete, but unrestricted interactions between workers and firms are allowed for, in which case the second step may be performed by maximum likelihood. We verify in simulations that these estimators perform well in data sets similar to the one of our application.

While clustering firms into classes in a first step makes the matched data problem tractable in our setup, two-step methods could also be useful in structural models. An attractive feature is that the classification does not rely on the entire model's structure, solely on the fact that unobserved firm heterogeneity operates at the class level. 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 estimates suggest 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 unobserved attributes. [TO BE COMPLETED]

Literature and outline. The methods we propose contribute to a large literature on the identification and estimation of models with latent heterogeneity. On the firm side, the k-means

[5] This result is derived under the assumption that the population of firms consists of a finite number of classes. In Bonhomme, Lamadon, and Manresa (2015) we consider a more general setting where the clustering is seen as approximating an underlying, possibly continuous, distribution of firm unobserved heterogeneity, and we provide consistency results and rates of convergence. However, in this alternative asymptotic framework, estimation error in the classification generally affects post-classification inference.

clustering algorithm we rely on is widely used in a number of fields, and efficient computational routines exist (Steinley, 2006). Discrete fixed-effects approaches have recently been proposed in single-agent panel data analysis (Hahn and Moon, 2010, Lin and Ng, 2012, Bonhomme and Manresa, 2015). Here we apply such an approach to models with two-sided heterogeneity.

On the worker side, we specify conditional mixture models, which allow distributions of worker types to depend on the conditioning firm classes. Nonparametric identification and estimation of finite mixtures have been extensively studied, see for example Hall and Zhou (2003), Henry, Kitamura, and Salanié (2014), Levine, Hunter, and Chauveau (2011), or Bonhomme, Jochmans, and Robin (2016). Identification of continuous mixture models is the subject of important work by Hu and Schennach (2008) and Hu and Shum (2012). Our conditional mixture approach is also related to mixed membership models, which have become popular in machine learning and statistics (Blei, Ng, and Jordan, 2003, Airoldi, Blei, Fienberg, and Xing, 2008).

Compared to previous work, we propose a hybrid approach that treats the firm classes as discrete fixed-effects and the worker types as (discrete or continuous) random-effects. This approach is motivated by the structure of typical matched employer employee data sets. 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.

Lastly, two recent innovative contributions rely on methods for models with latent heterogeneity to study questions related to worker/firm sorting on the labor market. Abowd, Schmutte, and McKinney (2015) propose a Bayesian approach where both firm and worker heterogeneity are discrete. Their setup allows for latent match effects to drive job mobility, in a way that is related to, but different from, our dynamic model. Unlike this paper they do not study identification formally, and they rely on joint (two-way random-effects) approach for estimation. Hagedorn, Law, and Manovskii (2014) propose to recover worker types by ranking workers by their earnings within firms, and aggregating those partial rankings across firms. Their method relies on long panels, and it exploits the implications of a specific structural model to identify firm heterogeneity.

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 empirical results. Lastly, we conclude in Section 7.

2 Framework of analysis

We consider an economy composed of N workers and J firms. We denote as j_{it} the identifier of the firm where worker i is employed at time t. Job mobility between a firm at t and another firm at $t+1$ is denoted as $m_{it} = 1$.

Firms are characterized by the *class* they belong to. We denote as k_{it} in $\{1, ..., K\}$ the class of firm j_{it}. The k's represent a partition of the set of firms into K classes, and k_{it} is a shorthand for $k(j_{it})$.[6] There could be as many classes as firms, in which case $K = J$ and $k_{it} = j_{it}$. Alternatively, firm classes could be defined in terms of observables such as industry or size. In Section 4 we will describe a method to consistently estimate the k_{it} from the data, under the assumption that they have K points of support.

Workers are also heterogeneous, and we denote the *type* of worker i as α_i. These types could be discrete or continuous, depending on the model specification. In addition to their unobserved types, workers may also differ in terms of their observable characteristics X_{it}.[7]

Lastly, worker i receives log-earnings Y_{it} at time t. The observed data for worker i is thus a sequence of earnings $(Y_{i1}, ..., Y_{iT})$, firm and mobility indicators $(j_{i1}, m_{i1}, ..., j_{i,T-1}, m_{i,T-1}, j_{iT})$, and covariates $(X_{i1}, ..., X_{iT})$. We consider a balanced panel setup for simplicity, and we focus on workers receiving positive earnings in each period.[8]

We consider two different models: a static model where current earnings do not affect job mobility or future earnings conditional on worker type and firm class, and a dynamic model that allows for these possibilities. We now describe these two models in turn. Next we discuss how our assumptions map to theoretical sorting models proposed in the literature. Throughout we denote $Z_i^t = (Z_{i1}, ..., Z_{it})$ the history of random variable Z_{it}.

2.1 Static model

Model and assumptions. In period 1, the type of a worker i, α_i, is drawn from a distribution that depends on the class k_{i1} of the firm where she is employed and her characteristics X_{i1}. The worker draws log-earnings Y_{i1} from a distribution that depends on α_i, k_{i1}, and X_{i1}.

At the end of every period $t \geq 1$, the worker moves to another firm (that is, $m_{it} = 1$ or 0) with a probability that may depend on her type α_i, her characteristics X_i^t, the fact that she

[6] In other words, $k : \{1, ..., J\} \mapsto \{1, ..., K\}$ maps firm j to firm class $k(j)$. In fact, classes could change over time, in which case k_{it} would be a shorthand for $k_t(j_{it})$.

[7] Here we abstract from firm-level observable characteristics. We return to this issue in Section 4.

[8] Incorporating an extensive employment margin within our framework could be done by adding a "non-employment" firm class $k_{it} = 0$, associated with $Y_{it} = \emptyset$.

moved in previous periods m_i^{t-1}, and current and past firm classes k_i^t. This probability, like all other probability distributions in the model, may depend on t unrestrictedly. Moreover, the probability that the class of the firm she moves to is $k_{i,t+1} = k'$ may also depend on α_i, X_i^t, m_i^{t-1}, and k_i^t. Lastly, covariates $X_{i,t+1}$ are drawn from a distribution depending on α_i, X_i^t, m_i^t, and k_i^{t+1}.

If the worker changes firm (that is, when $m_{it} = 1$), log-earnings $Y_{i,t+1}$ in period $t+1$ are drawn from a distribution that depends on α_i, $X_{i,t+1}$, and $k_{i,t+1}$. If instead the worker remains in the same firm between t and $t+1$ (that is, $m_{it} = 0$), $Y_{i,t+1}$ are drawn from an unrestricted distribution that may depend on Y_i^t, α_i, X_i^{t+1}, and k_i^{t+1}.

There are thus two main assumptions in the static model. First, job mobility may depend on the type of the worker and the classes of the firms, but not directly on earnings. As a result, the firm and mobility indicators, and firm classes, are all strictly exogenous in the panel data sense. In addition, covariates X_{it} are also strictly exogenous. Second, log-earnings after a job move may not depend on previous firm classes or previous earnings, conditional on the worker type and the new firm's class. Formally these two assumptions are the following.

Assumption 1.

(i) *(mobility determinants)* m_{it} *and* $k_{i,t+1}$ *are independent of* Y_i^t *conditional on* α_i, k_i^t, m_i^{t-1}, *and* X_i^t.

(ii) *(serial independence)* $Y_{i,t+1}$ *is independent of* Y_i^t, k_i^t, m_i^{t-1} *and* X_i^t *conditional on* α_i, $k_{i,t+1}$, $X_{i,t+1}$, *and* $m_{it} = 1$.

A simple example of the static model is the following log-earnings regression:

$$Y_{it} = a_t(k_{it}) + b_t(k_{it})\alpha_i + X_{it}' c_t + \varepsilon_{it}, \tag{1}$$

where $\mathbb{E}\left(\varepsilon_{it} \,|\, \alpha_i, k_i^T, X_i^T\right) = 0$. This model boils down to the one of Abowd, Kramarz, and Margolis (1999) in the absence of interaction effects, i.e. when $b_t(k) = 1$, and firms j_{it} and classes k_{it} coincide.[9]

Statistical implications on two periods. Here we consider the static model on $T = 2$ periods, which we will show to suffice for identification. Let $F_{k\alpha}(y_1)$ denote the cumulative distribution function (cdf) of log-earnings in period 1, in firm class k, for worker type α. Let

[9] While Assumption 1 (i) and the independence of future earnings on the previous firm class in Assumption 1 (ii) are instrumental to ensure identification, restrictions on the dependence structure of earnings are not needed to identify parameters such as $a_t(k)$, $b_t(k)$ and c_t in (1), as we will see below.

$F^m_{k'\alpha}(y_2)$ denote the cdf of log-earnings in period 2, for class k' and type α, for job movers between periods 1 and 2 (that is, when $m_{i1} = 1$). Let also $p_{kk'}(\alpha)$ denote the probability distribution of α_i for job movers between a firm of class k and another firm of class k'. Finally, let $q_k(\alpha)$ denote the distribution of α_i for workers in a firm of class k. All these distributions may be conditional on X_{i1} and X_{i2}, although we omit the conditioning for conciseness.

The model imposes the following restrictions on the bivariate log-earnings distribution for job movers:

$$\Pr\left[Y_{i1} \leq y_1, Y_{i2} \leq y_2 \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1\right] = \int F_{k\alpha}(y_1) F^m_{k'\alpha}(y_2) p_{kk'}(\alpha) d\alpha. \tag{2}$$

To see why (2) holds, note that Y_{i1} is independent of k_{i2} and m_{i1} conditional on α_i and k_{i1}. This is due to the fact that, by Assumption 1 (i), mobility is unaffected by log-earnings Y_{i1}, conditional on type and classes (and conditional on exogenous covariates). Moreover, Y_{i2} is independent of Y_{i1} and k_{i1}, conditional on α_i, k_{i2}, and $m_{i1} = 1$. This is due to the lack of dependence on the past after a job move in Assumption 1 (ii).

In addition, we have the following decomposition of the cdf of log-earnings in period 1:

$$\Pr\left[Y_{i1} \leq y_1 \,|\, k_{i1} = k\right] = \int F_{k\alpha}(y_1) q_k(\alpha) d\alpha. \tag{3}$$

Our main identification results under discrete or continuous worker heterogeneity, which we establish in Section 3, state that, under suitable rank conditions, $F_{k\alpha}$, $F^m_{k\alpha}$, and $p_{kk'}(\alpha)$, are identified from (2), and $q_k(\alpha)$ are then identified from (3). Moreover, we will show how to consistently estimate the partition of firms into classes based on the univariate or bivariate log-earnings distributions on the left-hand sides of (2) and (3).

The parameters in (2) and (3) allow documenting the sources of earnings inequality and the allocation of workers to firms. For example, the $F_{k\alpha}$ are informative about the presence of complementarities in the earnings function. Differences of $q_k(\alpha)$ across k are indicative of the cross-sectional sorting of high-earning workers to high-paying firms. Moreover, from the $p_{kk'}(\alpha)$ and data on transitions between classes one can recover estimates of the transition probabilities $\Pr\left(k_{i2} = k' \,|\, \alpha_i, k_{i1} = k, m_{i1} = 1\right)$ between types, which are informative about dynamic sorting patterns.

2.2 Dynamic model

Model and assumptions. There are two main differences between the dynamic model and the static one. First, at the end of period t the worker moves to another firm with a probability that depends on her current log-earnings Y_{it} in addition to her type α_i, X_{it}, and k_{it}, and likewise the probability to move to a firm of class $k_{i,t+1} = k'$ also depends on Y_{it}. A second difference is that log-earnings $Y_{i,t+1}$ in period $t+1$ are drawn from a distribution depending on the previous log-earnings Y_{it} and the previous firm class k_{it}, in addition to α_i, $X_{i,t+1}$, and $k_{i,t+1}$. Job movers and job stayers draw their log-earnings from different distributions conditional on these variables.[10] Formally we make the following assumptions.

Assumption 2.

(i) (mobility determinants) m_{it} and $k_{i,t+1}$ are independent of Y_i^{t-1}, k_i^{t-1}, m_i^{t-1} and X_i^{t-1} conditional on Y_{it}, α_i, k_{it}, and X_{it}.

(ii) (serial dependence) $Y_{i,t+1}$ is independent of Y_i^{t-1}, k_i^{t-1}, m_i^{t-1} and X_i^t conditional on Y_{it}, α_i, $k_{i,t+1}$, k_{it}, $X_{i,t+1}$, and m_{it}.

Assumption 2 consists of two first-order Markov conditions. In part (i), log-earnings Y_{it} are allowed to affect the probability to change job directly between t and $t+1$, but the previous earnings $Y_{i,t-1}$ do not have a direct effect. Similarly, in part (ii) log-earnings $Y_{i,t+1}$ may depend on the first lag of log-earnings Y_{it}, and on the current and lagged firm classes $k_{i,t+1}$ and k_{it}, but this dependence rules out effects of the further past such as $Y_{i,t-1}$ and $k_{i,t-1}$. Also note that, unlike in the static model, Assumption 2 (ii) restricts the evolution of log-earnings within as well as between jobs.

As a simple dynamic extension of (1) one may consider the following specification for the earnings of job movers between $t-1$ and t:

$$Y_{it} = \rho_t Y_{i,t-1} + a_{1t}(k_{it}) + a_{2t}(k_{i,t-1}) + b_t(k_{it})\alpha_i + X_{it}'c_t + \varepsilon_{it}, \tag{4}$$

where here log-earnings after a job move may depend on the values of earnings and firm class in the previous job.

Statistical implications on four periods. Here we describe the implications of the dynamic model on $T = 4$ periods. Let $G^f_{y_2,k\alpha}(y_1)$ (for "forward") denote the cdf of log-earnings in period 1, in a firm class k, for a worker of type α who does not change firm between periods 1 and 2 and earns y_2 in period 2. Let $G^b_{y_3,k'\alpha}(y_4)$ (for "backward") be the cdf of Y_{i4}, in firm class k', for a worker of type α who does not change firm between periods 3 and 4 and earns y_3 in period 3. Lastly, let $p_{y_2y_3,kk'}(\alpha)$ denote the type distribution of workers who stay in the same firm of

[10] In this model it is possible to let $X_{i,t+1}$ be drawn from a distribution that depends on Y_{it} as well as α_i, X_{it}, m_{it}, and $k_{i,t+1}$. Our identification arguments apply to this case, and estimation can readily be generalized to allow for predetermined individual characteristics, such as job tenure. In the empirical analysis we will consider only time-invariant covariates interacted with time effects, so we will not entertain this possibility.

class k between periods 1 and 2, move to another firm of class k' in period 3, remain in that firm in period 4, and earn y_2 and y_3 in periods 2 and 3, respectively.

The bivariate cdf of log-earnings Y_{i1} and Y_{i4} is, for workers who change firm between periods 2 and 3:

$$\Pr\left[Y_{i1} \leq y_1, Y_{i4} \leq y_4 \,|\, Y_{i2} = y_2, Y_{i3} = y_3, k_{i1} = k_{i2} = k, k_{i3} = k_{i4} = k', m_{i1} = 0, m_{i2} = 1, m_{i3} = 0\right] = \int G^f_{y_2,k\alpha}(y_1) G^b_{y_3,k'\alpha}(y_4) p_{y_2 y_3,kk'}(\alpha) d\alpha. \tag{5}$$

The reason why (5) holds is that, by Assumption 2 (i), Y_{i1} is independent of future mobility, firm classes, and earnings, conditional on Y_{i2}, k_{i1}, k_{i2}, and m_{i1}. Similarly, by Assumption 2 (ii), Y_{i4} is independent of past mobility, firm classes, and earnings, conditional on Y_{i3}, k_{i4}, k_{i3} and m_{i3}.

In addition, let $F_{k\alpha}$ be the cdf of log-earnings Y_{i2} for workers in firm class k who remain in the same firm in periods 1 and 2 (that is, $m_{i1} = 0$). Let also $q_k(\alpha)$ denote the distribution of α_i for these workers. The joint cdf of log-earnings in periods 1 and 2 is:

$$\Pr\left[Y_{i1} \leq y_1, Y_{i2} \leq y_2 \,|\, k_{i1} = k_{i2} = k, m_{i1} = 0\right] = \int G^f_{y_2,k\alpha}(y_1) F_{k\alpha}(y_2) q_k(\alpha) d\alpha. \tag{6}$$

The mathematical structure of (5) is analogous to that of (2). Intuitively, the conditioning on log-earnings Y_{i2} and Y_{i3} immediately before and after the job move ensures conditional independence of log-earnings Y_{i1} and Y_{i4}, although in this model earnings have a direct effect on job mobility and respond dynamically to lagged earnings and previous firm classes. Given the mathematical similarity between static and dynamic models, we will be able to extend the identification arguments to recover $G^f_{y_2,k\alpha}(y_1)$, $G^b_{y_3,k'\alpha}(y_4)$, and $p_{y_2 y_3,kk'}(\alpha)$ based on (5) using four periods data on workers moving between the second and third periods. Then, using only the first two periods we will show how to recover $F_{k\alpha}(y_2)$ and $q_k(\alpha)$ from (6).

2.3 Links with theoretical models

In this subsection we ask whether our assumptions are compatible with various theoretical models of the labor market. We consider models that abstract from labor supply, so we refer to earnings and wages indistinctively.

Relevant state space is (α, k_t). We first consider models where wages are a function, possibly non-linear or non-monotonic, of the worker type α, the firm class k_t, and a time-varying effect ε_t, where ε_t does not affect mobility decisions. This structure is compatible for instance with models where the wage paid to a worker does not have any history dependence, and where ε_t is classical measurement error or an i.i.d. match effect. In wage posting models (Burdett and Mortensen, 1998; Delacroix and Shi, 2006; Shimer, 2001)[11] workers climb a job ladder and their wages exhibit path dependence. However, since firms commit to posted wages, wages are independent of the worker history conditional on the firm class. This means that, while allowing for rich mobility and earnings patterns, these models are compatible with the assumptions of our static model, see Assumption 1.

Similarly, Assumption 1 is compatible with models where the wage is set as the outcome of a bargaining process between the firm and the worker, provided the worker's outside option is set to unemployment. This is the case in Shimer and Smith (2000) and Hagedorn, Law, and Manovskii (2014), where unemployment is a natural outside option since workers always go through unemployment before finding a new job. This is also the case of the model in Dey and Flinn (2005), where firms are not allowed to commit to future wages and bargain every period.

In sorting models such as Shimer and Smith (2000), specifying the wage function in a way that allows for interactions between worker types and firm classes is key. Indeed, in those models earnings may be non-monotonic in firm productivity, and workers rank identical firms differently. The static model can accommodate both features.

Markovian match effects and past firm. A natural extension is to allow workers to move based on the realization of the match effect ε_t, and to allow this match effect to be serially correlated. This is the case of our dynamic model under the assumption that ε_t is first-order Markov, see Assumption 2. The latter is compatible with a wage posting mechanism, or bargaining with value of unemployment, in the presence of Markovian match-specific effects. In those settings, a worker contemplating mobility compares the value at her current firm with her current match value, to the value at a new encountered firm with its corresponding match-specific draw. Alternatively, the Markovian ε_t may be thought of as a one-dimensional human capital accumulation process. This process can be deterministic or stochastic, as long as it is first-order Markov.

Our assumptions also allow for the past firm class to affect the wage after mobility, even beyond the past wage, see Assumption 2 (*ii*). In Postel-Vinay and Robin (2002), when an employed worker meets a new firm the two firms enter a Bertrand competition. The worker

[11] While the first two references do not explicitly allow for worker heterogeneity, one could pool segmented markets by worker types. Abowd, Kramarz, Pérez-Duarte, and Schmutte (2015), Shephard (2011), and Engbom and Moser (2015) have evaluated wage posting models empirically.

chooses to work in the firm with the highest surplus, and she extracts the surplus of the old firm. If the worker moves, her surplus and her earnings are functions of her type and both firms' classes. If she remains in the firm, the new wage realization depends on the worker type, the current firm's type, and the past wage, and is thus first-order Markov. In this setting, ε_t captures the current bargaining position of the worker, which is Markovian and (together with worker type and firm class) is a sufficient statistic for both the worker's and the firm's present values, and in mobility and new earnings' decisions. In addition, since mobility decisions are unrestricted beyond the first-order Markov assumptions, mobility can be the result of an endogenous search effort as in Bagger and Lentz (2014), where sorting happens because high-type workers benefit more from mobility and choose to search more intensively.

In contract posting models (Burdett and Coles, 2003; Shi, 2008) firms post wage-tenure contracts where payments are back-loaded to retain workers who are unable to commit to stay within the firm when offers come. In this environment, the optimal contract is Markovian conditional on worker type. The ε_t can be interpreted as the current level of utility promised to the worker by the firm, which in this model maps one-to-one into the current wage conditional on type. To nest these models it is crucial to allow the past firm to affect earnings after the move. This is due to the fact that the worker's value at the current firm depends on the firm class in addition of the current wage, so the past firm will affect the new wage beyond the level of the past wage.

Time effects. Our static and dynamic models allow distributions to depend unrestrictedly on calendar time. Lise and Robin (2013) develop a model of sorting in a labor market with sequential contracting and aggregate shocks. Present values and earnings are functions of worker and firm heterogeneity, as well as of an aggregate state and the current bargaining position. Assumption 2 of our dynamic model is satisfied in this setting. It is also satisfied in the model of Moscarini and Postel-Vinay (2009), where firms post wage contracts in an environment with aggregate shocks. Also, note that our assumptions allow for interactions between calendar time and unobserved worker type, thus allowing for deterministic and type-specific wage profiles.

Outside our framework. Any non-Markovian earnings structure will violate our assumptions. This will happen if the structural model allows for permanent plus transitory earnings dynamics conditional on worker types, as in Meghir and Pistaferri (2004) for example. This will also happen in models that combine a sequential contracting mechanism with a match-specific effect, as in Bagger, Fontaine, Postel-Vinay, and Robin (2011). In the latter case the agents need to keep track of both the match quality and the bargaining position, so we loose the one to one mapping between earnings and the value to the worker, making mobility decisions dependent on past wage conditional on the current wage, and thus violating our assumptions. Such environments are not nested in a framework that allows, in addition to observables X_{it}, for uni-dimensional time-varying effects ε_{it}.

Lastly, although it nests a number of theoretical models, without further assumptions our framework does not allow answering questions such as the efficiency or welfare consequences of the allocation of workers to jobs, and how they are affected by policy. Nevertheless, as they allow for complementarity and dynamics, the empirical methods we develop may help inform structural models. In Appendix D we evaluate the performance of one of our estimators to recover the contributions of worker and firm heterogeneity to earnings 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, and 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 to earnings dispersion and sorting patterns rather well, even under NAM.

3 Identification

In this section we show how, for a given partition of firms into classes and under suitable conditions, type-and-class-specific earnings distributions and class-specific type distributions are identified based on two periods in the static case, and four periods in the dynamic case. We first provide intuition in a simple version of the interactive regression model (1). Next we establish identification under an assumption that worker types are discretely distributed, without making further functional form assumptions. Finally we provide a nonparametric identification result when worker types are continuously distributed. The analysis in this section is conditional on a partition of firms into classes. In the next section we will show how to recover class membership $k_{it} = k(j_{it})$, for each firm j_{it}.

3.1 Intuition in an interactive regression model

To provide an intuition we start by considering a stationary specification of the interactive model of equation (1) with $T = 2$ periods, where neither $a_t(k)$ nor $b_t(k)$ depend on t, and we abstract from observed covariates X_{it}. Consider job movers between two firms of classes k and $k' \neq k$, respectively, between period 1 and 2. Here we study identification in a population where there is a continuum of workers moving between k and k'. This intuitively means that, in practice, this analysis will be relevant for data sets with a sufficient number of workers moving between firm classes. We will return to this issue in the estimation section, as this represents a motivation for our grouping of firms into classes. We have:

$$Y_{i1} = a(k) + b(k)\alpha_i + \varepsilon_{i1}, \quad Y_{i2} = a(k') + b(k')\alpha_i + \varepsilon_{i2}. \tag{7}$$

where $\mathbb{E}(\varepsilon_{it} \,|\, \alpha_i, k_{i1} = k, k_{i2} = k', m_{i1} = 1) = 0$. In this sample of job movers, the ratio $b(k')/b(k)$ is not identified without further assumptions.[12]

Consider now job movers from a firm in class k' to another firm in class k. We have:

$$Y_{i1} = a(k') + b(k')\alpha_i + \varepsilon_{i1}, \quad Y_{i2} = a(k) + b(k)\alpha_i + \varepsilon_{i2}.$$

It follows that:

$$\frac{b(k')}{b(k)} = \frac{\mathbb{E}_{kk'}(Y_{i2}) - \mathbb{E}_{k'k}(Y_{i1})}{\mathbb{E}_{kk'}(Y_{i1}) - \mathbb{E}_{k'k}(Y_{i2})}, \tag{8}$$

provided that the following holds:

$$\mathbb{E}_{kk'}(\alpha_i) \neq \mathbb{E}_{k'k}(\alpha_i), \tag{9}$$

where we have denoted $\mathbb{E}_{kk'}(Z_i) = \mathbb{E}(Z_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1)$. This shows that, if (9) holds, $b(k')/b(k)$ is identified from mean restrictions.

An intuition for (8) is that, by comparing differences in log-earnings between two different subpopulations of workers in firm class k' and in class k, the ratio is informative about the effects of worker heterogeneity in the two firm classes. This requires the types of workers moving from k to k' and from k' to k to differ. Note that, if $b(k') + b(k) \neq 0$, the latter condition is equivalent to:

$$\mathbb{E}_{kk'}\left(Y_{i1} + Y_{i2}\right) \neq \mathbb{E}_{k'k}\left(Y_{i1} + Y_{i2}\right), \tag{10}$$

so it can be empirically tested.

The mean restrictions in (8) hold irrespective of the serial dependence properties of ε_{it}. In addition, these restrictions are linear in parameters (in this case, the $b(k)'s$). In Appendix C we show that both features are preserved in more general static and dynamic interactive regression models such as (1) and (4). In interactive models it is thus possible to relax the assumption that the log-earnings of job movers are serially independent conditional on worker

[12] Model (7) 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 identification fails in general. For example, $b(k')/b(k)$ is not identified when ε_{i1}, ε_{i2}, and α_i are independent Gaussian random variables (Reiersøl, 1950).

type and firm classes, and still identify the a's, b's, and means of α_i. In contrast, to identify the within-firm-class variances of worker types in finite-length panels, restrictions must be imposed on the dependence structure of the ε's, such as independence between ε_{i1} and ε_{i2} when $T = 2$. In Appendix C we describe these interactive models in detail, and provide conditions for identification.

3.2 Identification results

In this subsection we consider general static and dynamic models under Assumptions 1 and 2, respectively. We make no functional form assumptions on earnings distributions, except that we consider models where worker types α_i have finite support. Relying on discrete types is helpful for tractability, and we will use a finite mixture specification in our empirical implementation. However, at the end of this subsection we also provide an identification result for continuously distributed worker types.

We start by considering the static model on two periods. The dynamic model having a similar mathematical structure as the static one, the identification arguments are closely related (see below). Let L be the number of points of support of worker types, and let us denote the types as $\alpha_i \in \{1, ..., L\}$. We assume that L is known.[13] All distributions below may be conditional on (X_{i1}, X_{i2}), although we omit the conditioning for conciseness.[14]

In this finite mixture model, (2) and (3) imply restrictions on the cdfs $F_{k\alpha}$ and $F^m_{k'\alpha}$, and on the probabilities $p_{kk'}(\alpha)$ and $q_k(\alpha)$. We now show that these objects are all identified subject to suitable conditions. For this we start with a definition.

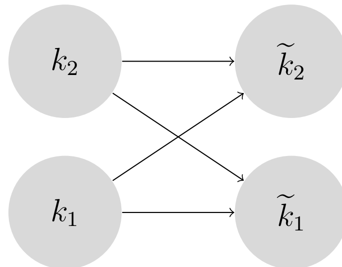
Definition 1. *An alternating cycle of length R is a pair of sequences of firm classes $(k_1, ..., k_R)$ and $(\widetilde{k}_1, ..., \widetilde{k}_R)$, with $k_{R+1} = k_1$, such that $p_{k_r,\widetilde{k}_r}(\alpha) \neq 0$ and $p_{k_{r+1},\widetilde{k}_r}(\alpha) \neq 0$ for all r in $\{1, ..., R\}$ and α in $\{1, ..., L\}$.*

Assumption 3. *(mixture model)*
(i) For any two firm classes $k \neq k'$ in $\{1, ..., K\}$, there exists an alternating cycle $(k_1, ..., k_R)$, $(\widetilde{k}_1, ..., \widetilde{k}_R)$, such that $k_1 = k$ and $k_r = k'$ for some r, and such that the scalars $a(1), ..., a(L)$

[13] Identifying and estimating the number of types in finite mixture models is a difficult question. Kasahara and Shimotsu (2014) provide a method to consistently estimate a lower bound.

[14] Specifically, with time-varying covariates the identification argument goes through provided $F_{k\alpha x_1}$ and $F^m_{k'\alpha x_2}$ solely depend on period-specific covariates. With time-invariant covariates it is not possible to link type probabilities across covariates values, due to a labelling problem. This issue echoes the impossibility to identify the coefficients of time-invariant regressors in fixed-effects panel data regressions.

Figure 1: An alternating cycle of length $R = 2$



are all distinct, where:

$$a(\alpha) = \frac{p_{k_1,\widetilde{k}_1}(\alpha)p_{k_2,\widetilde{k}_2}(\alpha)...p_{k_R,\widetilde{k}_R}(\alpha)}{p_{k_2,\widetilde{k}_1}(\alpha)p_{k_3,\widetilde{k}_2}(\alpha)...p_{k_1,\widetilde{k}_R}(\alpha)}.$$

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

(ii) For a suitable finite set of values for y_1 and y_2, which includes $(+\infty, +\infty)$, and for all r in $\{1, ..., R\}$, the matrices $A(k_r, \widetilde{k}_r)$ and $A(k_r, \widetilde{k}_{r+1})$ have rank K, where:

$$A(k, k') = \{\Pr[Y_{i1} \le y_1, Y_{i2} \le y_2 \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1]\}_{(y_1,y_2)}.$$

Assumption 3 requires that any two firm classes k and k' belong to an alternating cycle. An example is given in Figure 1, in which case the presence of an alternating cycle requires that there is a positive proportion of every worker type in the sets of movers from k_1 to $\widetilde{k}_1$, k_1 to $\widetilde{k}_2$, k_2 to $\widetilde{k}_1$, and k_2 to $\widetilde{k}_2$, respectively. 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. Another difference is that in our context the relevant notion of connectedness is between firm classes, as opposed to between individual firms.

Assumption 3 (i) requires some asymmetry in worker type composition between different firm classes. This condition requires either non-random cross-sectional sorting or non-random mobility, as it fails when $p_{kk'}(\alpha)$ does not depend on (k, k'). Another case where part (i) fails is when cross-sectional sorting and sorting associated with job mobility exactly offset each other, so $p_{kk'}(\alpha)$ is symmetric in (k, k'). 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 D. In the mixture model analyzed here, the presence of asymmetric job movements between firm classes is crucial for identification. This is similar to the case of the simple interactive model studied above, see (10). In the empirical analysis we will provide evidence of such asymmetry. The requirements on cycles can be relaxed, at the cost of loosing identification of some of the quantities of interest. In Appendix B we illustrate this point in a model where worker types and firm classes are ordered, there is strong positive assortative matching, and workers only move between "nearby" firm classes.

Assumption 3 (*ii*) is a rank condition. It will be satisfied if, in addition to part *i*), for all r the distributions $F_{k_r,1}$, ..., $F_{k_r,L}$ are linearly independent, and similarly for $F_{\widetilde{k}_r,1}$, ..., $F_{\widetilde{k}_r,L}$, $F^m_{k_r,1}$, ..., $F^m_{k_r,L}$, and $F^m_{\widetilde{k}_r,1}$, ..., $F^m_{\widetilde{k}_r,L}$. Note that this assumption is in principle testable.

The next result shows that, with only two periods and given the structure of the static model, both the type-and-class-specific earnings distributions and the proportions of worker types for job movers can be uniquely recovered. The intuition for the result is similar to that in the simple interactive regression model above. Due to the discrete heterogeneity setting, identification holds up to a choice of labelling of the latent worker types. All proofs are in Appendix A.

Theorem 1. *Let $T = 2$, and consider the joint distribution of log-earnings of job movers. Let Assumptions 1 and 3 hold. Suppose that firm classes k_{it} are observed. Then, up to common labelling of the types α, $F_{k\alpha}$ and $F^m_{k'\alpha}$ are identified for all (α, k, k'). Moreover, for all pairs (k, k') for which there are job moves from k to k', $p_{kk'}(\alpha)$ is identified for all α, up to the same labelling.*

The next corollary shows that the proportions of worker types α in each firm class k in period 1 are also identified.

Corollary 1. *Let $T = 2$. Consider the distribution of log-earnings in the first period. Let Assumptions 1 and 3 hold. Suppose that firm classes k_{it} are observed. Then the type proportions $q_k(\alpha)$ are identified up to the same labelling as in Theorem 1.*

Dynamic model. A similar approach allows us to establish identification of the dynamic mixture model on four periods with discrete worker heterogeneity. Exploiting the link between (2) and (5) on the one hand, and (3) and (6) on the other hand, we obtain the following corollary to Theorem 1. The required assumptions, particularly on the existence of cycles, are more stringent than in the static case, see Appendix A.[15]

[15]Identification here relies on within-job variation in earnings. Alternatively, one could rely on multiple job moves per worker.

Corollary 2. *Let $T = 4$. Consider the joint distribution of log-earnings of job movers. Let Assumption 2 hold. Let also Assumption 3 hold, with Y_{i2} replaced by Y_{i4}, k replaced by (k, y_2), and k' replaced by (k', y_3); see Appendix A for a precise formulation. Suppose that firm classes k_{it} are observed. Then, up to common labelling of the types α:*

(i) $G^f_{y_2,k\alpha}$ and $G^b_{y_3,k'\alpha}$ are identified for all (α, k, k'). Moreover, for all (k, y_2, k', y_3) for which there are job moves from (k, y_2) to (k', y_3), $p_{y_2y_3,kk'}(\alpha)$ is identified for all α.

(ii) $F_{k\alpha}$ and $q_k(\alpha)$, and log-earnings cdfs in periods 3 and 4, are also identified. Lastly, transition probabilities between firm classes are identified.

Continuous worker types. Let worker types α_i be continuously distributed. Here we focus on the static model, but similar arguments apply to the dynamic model. As in Hu and Schennach (2008) (HS hereafter) we assume bounded joint and conditional densities. We have, by Assumption 1 and for all k, k':

$$f_{kk'}(y_1, y_2) = \int f_{k\alpha}(y_1) f^m_{k'\alpha}(y_2) p_{kk'}(\alpha) d\alpha, \tag{11}$$

where the f's are densities corresponding to the cdfs in (2).

The structure of (11) is related to that in HS. Indeed, Assumption 1 implies that Y_{i1} and Y_{i2} are independent conditional on $\alpha_i, k_{i1}, k_{i2}, m_{i2} = 1$. However, here independence holds between two outcomes only, while HS assume conditional independence between three outcomes. Nevertheless, under conditions related to those in HS, by relying in addition on the network structure of workers' movements between firms it is possible to establish nonparametric identification using similar arguments as in the proof of Theorem 1.

Specifically, in Appendix B we formally show identification of earnings and type distributions using an alternating cycle of length $R = 2$. There, we show that the model's assumptions imply restrictions on certain operators that mimic the matrix restrictions appearing in the proof of Theorem 1. Identification then relies on nonparametric analogs to the rank conditions in Assumption 3 (ii), also called *completeness* conditions. In particular, these conditions require α_i to be one-dimensional.

In addition, in order to identify the log-earnings densities $f_{k\alpha}$ and $f^m_{k'\alpha}$, the conditions we borrow from HS and Hu and Shum (2012) involve a monotonicity assumption. For example, the latter is satisfied if $\mathbb{E}[Y_{i1} \mid \alpha_i = \alpha, k_{i1} = k]$ is monotone in α. This condition might be natural if α represents a worker's productivity type, for example, although it rules out non-monotonic earnings profiles and multi-dimensional worker types. Note in contrast that, when worker types are assumed to have a known finite support (as in Theorem 1), no such assumption is needed and the only ambiguity lies in the arbitrary labelling of the latent types.

4 Estimation

The results in the previous section show that, provided that different types of workers make transitions between firm classes, earnings distributions can be identified in the presence of sorting and complementarities. These results hold at the firm class level k_{it}, where in principle the k_{it} could coincide with the firm j_{it}. However, in matched employer-employee panel data sets of typical sizes, estimating models with worker/firm interactions, dynamics, and possibly flexible distributional forms for earnings may be impractical due to the incidental parameter biases caused by the large number of firm-specific parameters. For this reason, we introduce a dimension-reduction method to partition firms into firm classes, which is consistent with our static or dynamic models of earnings and mobility. Then we describe a two-step estimation approach that takes estimated firm classes as inputs, and recovers earnings and mobility parameters in a second stage.

4.1 Recovering firm classes using clustering

Clustering earnings distributions. In both the static and dynamic models described in Section 2, the distributions of log-earnings Y_{it} and characteristics X_{it}, and the probabilities of mobility m_{it}, are all allowed to depend on firm classes k, but not on the identity of the firm within class k. In other words, unobservable firm heterogeneity operates at the level of firm classes in the model, not at the level of individual firms. This key observation motivates classifying firms into classes in terms of their distributions of observables, as we now explain.

For example, in (3) the first period's distribution of log-earnings in firm j does not depend on j beyond its dependence on firm class $k = k(j)$:

$$\Pr\left[Y_{i1} \leq y_1 \,|\, j_{i1} = j\right] \;=\; \int F_{k\alpha}(y_1) q_k(\alpha) d\alpha, \tag{12}$$

where the left-hand side thus only depends on $k = k(j)$.

Motivated by (12) we propose partitioning the J firms in the sample into classes by solving the following weighted k-means problem:

$$\min_{k(1),\ldots,k(J),H_1,\ldots,H_K} \sum_{j=1}^{J} n_j \int \left(\widehat{F}_j(y) - H_{k(j)}\left(y\right)\right)^2 d\mu(y), \tag{13}$$

where $\widehat{F}_j$ denotes the empirical cdf of log-earnings in firm j, n_j is the number of workers in firm j (both in period 1), and μ is a discrete or continuous measure. The minimization in (13) is with respect to all possible partitions of the J firms into K groups (the $k(j)$'s), and to class-specific cdfs (the H_k's). Finding global minima is often challenging. However, k-means algorithms are widely used in many fields, and efficient heuristic computational methods have been developed (e.g., Steinley, 2006).

To provide a formal justification for the classification in (13), we consider a setting where the model (either static or dynamic) is well-specified and there exists a partition of the J firms into K classes in the population. We consider an asymptotic sequence where the number of firms J may grow with the number of workers N and the numbers of workers per firm n_j. We make the following assumptions, where we take the measure μ to be discrete on $\{y_1, ..., y_D\}$, $k^0(j)$ denote firm classes in the population, H_k^0 denote the population class-specific cdfs, and $\|H\|^2 = \sum_{d=1}^{D} H(y_d)^2$.

Assumption 4. *(clustering)*

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

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

(iii) For all $k \neq k'$ *in* $\{1, ..., K\}$, $\|H_k^0 - H_{k'}^0\| > 0$.

(iv) Let $n = \min_{j=1,...,J} n_j$. *There exists* $\delta > 0$ *such that* $J/n^\delta \to 0$ *as* n *tends to infinity.*

Assumption 4 (i) could be relaxed to allow for weak dependence both across and within firms, in the spirit of the analysis of Bonhomme and Manresa (2015) who focused on a panel data context, i.e. they analyzed data on individuals over time as opposed to workers within firms. Parts 4 (ii) and (iii) require that the clusters be large and well-separated in the population. We further discuss part (iii) below. Assumption 4 (iv) allows for asymptotic sequences where the number of workers per firm grows polynomially more slowly than the number of firms. Note that while it imposes conditions on the rate of growth of the minimum firm size, this condition allows some firms to asymptotically represent a non-vanishing fraction of the sample.

Verifying the assumptions of Theorems 1 and 2 in Bonhomme and Manresa (2015), we now show that the estimated firm classes, $\widehat{k}(j)$, converge uniformly to the population ones up to an arbitrary labelling. As a result, we obtain that the asymptotic distribution of the log-earnings cdf $\widehat{H}_k$ coincides with that of the empirical cdf of wages in the population class k (that is, the true one). In practice this means that, provided Assumption 4 holds, one can treat the estimated firm classes as known when computing standard errors of estimators based on them. Here we prove a pointwise result for $\widehat{H}_k$, but the equivalence also holds uniformly in y. It also holds for second-step estimates based on pre-estimated classes (see the next subsection).

Proposition 1. *Let Assumption 4 hold. Then, up to labelling of the classes k:*

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

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

Alternative clustering approaches. In practice, instead of cdfs one could cluster features of earnings distributions such as means, variances, or other moments of the firm-specific distributions. One could also cluster more general firm-specific distributions in addition to cross-sectional cdfs of earnings. For example, under the assumption that they are class-specific one could add firm size, industry, output, profit or value added, as additional measurements to the k-means classification (13). Proposition 1 then provides an asymptotic justification for the method in settings where there is a finite number of latent firm classes in the population.[16]

A particular difficulty with identifying firm classes from cross-sectional observations only is that it might be that two cross-sectional earnings distributions coincide between two firms, one offering a higher earnings schedule but having low-type workers, the other one offering a lower earnings schedule but having high-type workers. This possibility has been emphasized in the theoretical sorting literature (e.g., Eeckhout and Kircher, 2011). It is reflected in the violation of the separation condition in Assumption 4 *(iii)*. So it may be impossible to separate two different classes from the cross-section, even though their conditional earnings distributions given worker types are different.

A different classification approach, which has the potential to address this issue, is to cluster firms based on multivariate distributions of earnings in the panel. To see how this can be done, consider the static model on two periods. We have, by (2) and due to the class-specific nature of firm heterogeneity:

$$\Pr\left[Y_{i1}\leq y_1,Y_{i2}\leq y_2\,|\,j_{i1}=j,j_{i2}=j',m_{i2}=1\right] \;=\; \int F_{k\alpha}(y_1)F_{k'\alpha}^m(y_2)p_{kk'}(\alpha)d\alpha, \qquad (14)$$

which does not depend on (j,j') beyond its dependence on $k=k(j)$ and $k'=k(j')$.

This motivates the following bi-clustering method to classify firms into classes:

$$\min_{k(1),\ldots,k(J),G_{11},\ldots,G_{KK}}\sum_{i=1}^{N_m}\int\int\left(\mathbf{1}\{Y_{i1}\leq y_1\}\mathbf{1}\{Y_{i2}\leq y_2\}-G_{k(j_{i1}),k(j_{i2})}\left(y_1,y_2\right)\right)^2 d\mu(y_1,y_2), \qquad (15)$$

for a bivariate measure μ, where the first N_m individuals in the sample are the job movers between periods 1 and 2. Algorithms to solve (15) have been comparatively less studied than

[16] In Bonhomme, Lamadon, and Manresa (2015) we study asymptotic properties of k-means clustering and two-step methods in settings where the K clusters are viewed as approximating a possibly continuous heterogeneity structure.

k-means classification problems such as (13). At the same time, as we show in Appendix B in the static mixture model with discrete types, the separation condition for consistency of classification in (15) is weaker than in the cross-sectional case of Assumption 4 (*iii*).[17]

Finally, note that the clustering method in (13) estimates firm classes as "discrete fixed-effects", allowing them to be correlated to firm-specific covariates. In our application on short panels we will assume that the firms' classification is time-invariant, and correlate the estimated classes *ex-post* to firm observables. In longer panels, the clustering method could be generalized to account for time-varying classes, and one could document how the evolution of the classes relates to time variation in observables such as firm size, as we outline in Appendix B.

4.2 Two-step estimation

Our estimation strategy consists of two steps. In the first step we compute estimated firm classes, $\widehat{k}(j)$, for all firms j in the sample, by solving a classification problem such as (13). In the second step we impute a class $\widehat{k}_{it} = \widehat{k}(j_{it})$ to each worker-period observation in the sample, and we estimate the model conditional on the $\widehat{k}_{it}$'s.[18]

The second step takes a different form, depending on the model considered (either static or dynamic, interactive-based or mixture-based). Importantly, the classification step does not rely on the model's structure. In fact one could use a similar two-step approach in structural settings, and estimate the structural model in a second step given the estimated $\widehat{k}_{it}$'s. In that case too, one would not need to impose the model's structure in order to perform the classification in the first step.

Consider a finite mixture specification of the static model on two periods. One possibility, which we use in the empirical analysis, is to use a parametric model such as a Gaussian mixture model with (k, α)-specific wage means and variances. We also experimented with a specification where $F_{k\alpha}$ follows a mixture-of-normals with (k, α)-specific parameters. The log-likelihood conditional on estimated firm classes takes the form:

$$\sum_{i=1}^{N_m} \sum_{k=1}^{K} \sum_{k'=1}^{K} \mathbf{1}\{\widehat{k}_{i1} = k\} \mathbf{1}\{\widehat{k}_{i2} = k'\} \ln \left(\sum_{\alpha=1}^{L} p_{kk'}(\alpha) f_{k\alpha}(Y_{i1}; \theta) f^m_{k'\alpha}(Y_{i2}; \theta) \right), \tag{16}$$

where $f_{k\alpha}(y; \theta)$ and $f^m_{k\alpha}(y; \theta)$ are parametric log-earnings densities indexed by θ, and the pro-

[17] One could add job stayers in (15). Also, the two objective functions in (13) and (15) could be combined, thus incorporating both cross-sectional and longitudinal information into the classification.

[18] As a robustness check, in order to avoid biases due to overfitting (Abadie, Chingos, and West, 2014), we also use the following split-sample method: estimate firm classes using job stayers only, and then estimate log-earnings distributions using job movers.

portions $p_{kk'}(\alpha)$ 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 (2016) and Levine, Hunter, and Chauveau (2011). These methods could also be used in the present context.

Given estimates of the firm classes and the cdfs $F_{k\alpha}$ we estimate the type proportions based on another, simpler, finite mixture problem. The first period's log-likelihood is, given estimated firm classes and distributions:

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{1}\{\widehat{k}_{i1}=k\}\ln\left(\sum_{\alpha=1}^{L}q_{k,X_i}(\alpha)f_{k\alpha}(Y_{i1};\widehat{\theta})\right), \tag{17}$$

where $q_{kx}(\alpha)$ denotes the proportion of type α workers in class k with covariate x. In practice we use a second EM algorithm to maximize (17).[19]

We estimate the proportions $q_{kx}(\alpha)$ in (17) by allowing them to depend on time-invariant worker covariates, X_i, such as education or age in the first period. This specification allows us to distinguish sorting in terms of x from sorting in terms of unobservables. For example, for all (k,α) we can write:

$$q_k(\alpha) \;=\; \sum_x p(x)q_{kx}(\alpha) + \sum_x \left(p_k(x)-p(x)\right)q_{kx}(\alpha), \tag{18}$$

where $p_x=\Pr(X_i=x)$, and $p_k(x)=\Pr(X_i=x\,|\,k_{i1}=k)$. The first term on the right-hand side of (18), say $\widetilde{q}_k(\alpha)$, 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 (18) reflect the contribution of unobservables and observables, respectively, to differences in worker type composition across firm classes. Note that one also could introduce observable characteristics in $f_{k\alpha}$, $f_{k\alpha}^m$, and $p_{kk'}(\alpha)$ in (16). Given the short length of the panel, in the empirical analysis we will use a fully nonstationary specification, and account for time-invariant covariates using (17).

We use a similar approach to estimate the dynamic finite mixture model on four periods, based on two EM algorithms to estimate the earnings and type distributions and the mobility probabilities. Specifically, we let the conditional mean of Y_{i4} given Y_{i3} and worker and firm classes be $\mu_{4k'\alpha}+\rho_{4|3}Y_{i3}$, where $\mu_{4k\alpha}$ is a (k,α)-specific intercept. Likewise, the conditional

[19] Type proportions could alternatively be estimated using a least squares regression, as can be seen from the identification arguments in Section 3.

mean of Y_{i1} given Y_{i2} and worker and firm classes is $\mu_{1k\alpha} + \rho_{1|2}Y_{i2}$. The joint distribution of (Y_{i2}, Y_{i3}) for job movers between classes k and k' has means $(\mu_{2k\alpha} + \xi_2(k'), \mu_{3k'\alpha} + \xi_3(k))$, and (k, k')-specific covariance matrix. We denote its density as $f^m_{kk'\alpha}(y_2, y_3)$. The distribution of Y_{i2} for workers in class k has (k, α)-specific mean and k-specific variance. The two likelihood maximizations are based on:

$$\sum_{i=1}^{N_m} \sum_{k=1}^{K} \sum_{k'=1}^{K} \mathbf{1}\{\widehat{k}_{i2} = k\}\mathbf{1}\{\widehat{k}_{i3} = k'\} \ln \left(\sum_{\alpha=1}^{L} p_{kk'}(\alpha) f^f_{Y_{i2},k\alpha}(Y_{i1}; \theta) f^b_{Y_{i3},k'\alpha}(Y_{i4}; \theta) f^m_{kk'\alpha}(Y_{i2}, Y_{i3}; \theta) \right), \tag{19}$$

and:

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \mathbf{1}\{\widehat{k}_{i2} = k\} \ln \left(\sum_{\alpha=1}^{L} q_{k,X_i}(\alpha) f^f_{Y_{i2}k\alpha}(Y_{i1}; \widehat{\theta}) f_{k\alpha}(Y_{i2}; \nu) \right), \tag{20}$$

where the second maximization delivers estimates of ν and $q_{kx}(\alpha)$. The parameters $\rho_{1|2}$, ρ_{23} and $\rho_{4|3}$ are included in θ.[20]

This approach can handle models with discrete worker heterogeneity in general settings. In addition, although we do not pursue this route here one could use a similar mixture-based approach to estimate models with continuously distributed α's. When additional structure is imposed on means and covariances of earnings, estimation can be based on simple restrictions involving these moments. This is the case in the AKM model, and in interactive models such as (1) and (4). As we show in Appendix C, two-step methods deliver computationally convenient estimation algorithms in static and dynamic interactive regression models.

Lastly, in this section we have focused on short panels. In the empirical application we will report estimation results based on two and four periods, relying on both job movers and job stayers. In Appendix B we describe estimation on T periods.

Additional steps. Once the model's parameters have been estimated, one can re-classify firms into classes. As an example, given estimates $\widehat{k}_{i2}$ (from the first step) and $\widehat{G}_{kk'}$ (from the second step) one can re-classify firms in period 1 as follows:

$$\min_{k(1),\ldots,k(J)} \sum_{i=1}^{N_m} \int\int \left(\mathbf{1}\{Y_{i1} \leq y_1\}\mathbf{1}\{Y_{i2} \leq y_2\} - \widehat{G}_{k(j_{i1}),\widehat{k}_{i2}}(y_1, y_2) \right)^2 d\mu(y_1, y_2), \tag{21}$$

and compare the estimates $\widehat{k}(j_{i1})$ with the original $\widehat{k}_{i1}$.

[20] Note that one could alternatively maximize (19) and (20) jointly, and add periods 3 and 4 outcomes to (20).

Incidental parameters: remark. The estimators in this section involve a fairly large number of parameters. For example, type proportions may be poorly estimated for workers moving between classes k and k', when the total number of workers making this transition, say $N^m_{kk'}$, is small. In practice interest often centers on class-specific parameters such as type-and-class-specific mean log-earnings, or type proportions in a firm class. A simple approach is then to trim out (k, k') cells when $N_{kk'}$ is smaller than a threshold (e.g., 20 or 50). Alternatively, one could treat (k, k')-specific parameters as "random-effects" and integrate them out using a prior distribution. The latter approach is widely used in text analysis and machine learning, see for example Blei, Ng, and Jordan (2003). We have experimented with both approaches and found minor effects on the empirical results.

5 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. The Swedish data registry (ANST), which is part of the register-based labor market statistics at Statistics Sweden (RAMS), provides information about individuals, their employment, and their employers. This database is collected yearly from the firm's income statements. The other databases provide additional information on worker and firm characteristics, as well as unemployment status of workers: LOUISE (or LINDA) contains information on the workers, SBS provides accounting data and balance sheet information for all non-financial corporations in Sweden, and the Unemployment Register contains spells of unemployment registered at the Public Employment Service.

The RAMS dataset allows constructing individual employment spells within a year, as it provides the first and last remunerated month for each employee in a plant as well as firm and plant identifier. In this paper we define firms through firm identifiers. We define the main employment of each individual in a year as the one providing the highest earnings in that year. The main employer determines the employer of a worker in a given year. RAMS provides pre-tax yearly earnings for each spell. We use the ratio between total earnings at the main employer and the number of months employed as our measure of monthly earnings. We compute real earnings in 2007 prices.

Sample selection. 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. In addition, all workers from the public sector or self-employed are discarded.

We focus on workers employed in years 2002 and 2004. These two years correspond to periods 1 and 2 in the static model. We restrict the sample to males. We choose not to include female workers in the analysis in order to avoid dealing with gender differences in labor supply, since we do not have information on hours worked. We keep firms which have at least one worker who is fully employed in both 2002 and 2004 ("continuing firms"), where fully employed workers are those employed in all twelve months in a year in one firm. From this 2002-2004 sample we define the sub-sample of movers as workers whose firm identifier changes between 2002 and 2004.[21]

Restricting workers to be fully employed in 2002 and 2004, and firms to be present in both periods, is not innocuous, and we will see that this results in a substantial reduction of the number of workers whose firm identifier changes in the course of 2003. The reason for this conservative sample selection is that we want to capture, as closely as possible, individual job moves between existing firms. In particular, a firm may appear in only one period because of a merger or acquisition, entry or exit, or due to a re-definition of the firm identifier over time. In our preferred specification we do not include these job moves as we do not think that they map naturally to our model.

For the dynamic model we consider a subsample that covers the years 2001 to 2005. In addition to the criteria used to construct the 2002-2004 sample, we require that workers be fully employed in the same firm in 2001 and 2002, and in 2004 and 2005.

Descriptive Statistics We now report descriptive statistics on the 2002-2004 and 2001-2005 samples, as well as on the subsamples of job movers. Results can be found in Table 1.

The 2002-2004 sample contains about 600,000 workers and 44,000 firms. Hence the average number of workers per firm is 13.7. The mean firm size as reported by the firm is higher, 37.6, due to our sample selection that focuses on fully employed male workers. In the 2001-2005 sample, the mean number of workers and mean reported size are 12.3 and 37.1, respectively. The distribution of firm size is skewed, and medians are smaller. At the same time, reported firm sizes in the subsamples of movers are substantially higher. This implies that classification accuracy, in the first step of the algorithm, will be higher for those firms, who are key in order to estimate earnings distributions. In the next section we will report the results of several

[21] If a worker returns in 2004 to the firm he worked for in 2002 we do not consider this worker to be a mover. This represents 4.3% of the 2002-2004 sample.

Table 1: Data description

years:	2002-2004 all	2002-2004 movers	2001-2005 all	2001-2005 movers
number of workers	599,775	19,557	442,757	9,645
number of firms	43,826	7,557	36,928	4,248
number of firms ≥ 10	23,389	6,231	20,557	3,644
number of firms ≥ 50	4,338	2,563	3,951	1,757
mean firm reported size	37.59	132.33	39.67	184.77
median firm reported size	10	28	11	36
firm reported size for median worker	154	159	162	176
firm actual size for median worker	72	5	64	3
% high school drop out	20.6%	14%	21.5%	14.7%
% high school graduates	56.7%	57.3%	57%	59%
% some college	22.7%	28.7%	21.4%	26.3%
% workers younger than 30	16.8%	28%	13.9%	23.8%
% workers between 31 and 50	57.2%	59%	59.4%	62.1%
% workers older than 51	26%	13%	26.7%	14.2%
% workers in manufacturing	45.4%	35.1%	48.5%	40.4%
% workers in services	25.3%	33.7%	22.4%	27.8%
% workers in retail and trade	16.7%	20.3%	16.3%	20.8%
% workers in construction	12.6%	10.9%	12.8%	11%
mean log-earnings	10.18	10.17	10.19	10.17
variance of log-earnings	0.124	0.166	0.113	0.148
between-firm variance of log-earnings	0.0475	0.1026	0.0441	0.0947
mean log-value-added per worker	15.3	16.35	15.37	16.63

Notes: Swedish registry data. Males, fully employed in the same firm in 2002 and 2004 (columns 1 and 2), and fully employed in the same firm in 2001-2002 and 2004-2005 (columns 3 and 4), continuously existing firms. Figures for 2002. Mean log value-added per worker reported for firms with positive value-added (98.7% of firms in the 2002-2004 sample).

simulation exercises aimed at assessing the accuracy of our classification given this distribution of firm size.

Identification relies on workers moving between firms over time. In the 2002-2004 sample, the mobility rate, which we define as the fraction of workers fully employed in 2002 and 2004 whose firm identifiers are different in these two years, is $19557/599775 = 3.3\%$. In the 2001-2005 sample the rate is 2.4%. These numbers are lower than the ones calculated by Skans, Edin, and Holmlund (2009), who document between-plant mobility rates ranging between 4% and 6% between 1986 and 2000.[22] To understand how our sample selection influences the mobility rate, we have computed similar descriptive statistics on the entire 2002-2004 sample, without imposing that workers are fully employed in 2002 and 2004 or that firms exist in the two periods, see Table E2 in Appendix E. Removing the requirements of full-year employment in both 2002 and 2004 and continuously existing firms results in a considerably less restrictive definition of mobility, as the mobility rate is 11.2% in this case.[23] Although we prefer to focus on a more restrictive definition in the baseline estimation, as a robustness check we will report parameter estimates on this larger sample.

The between-firm log-earnings variance represents 38.3% of total log-earnings variance in 2002. This number is higher than the 31% percentage explained between plants in 2000, as reported by Skans, Edin, and Holmlund (2009). However, despite growing steadily over the past decades, the between-firm (or plant) component is still lower compared with other economies such as Germany, Brazil, or the US. In Germany and Brazil, between components are closer to 50%, see Baumgarten, Felbermayr, and Lehwald (2014) or Akerman, Helpman, Itskhoki, Muendler, and Redding (2013), for example. In the US, Barth, Bryson, Davis, and Freeman (2014) report a between-establishment log-earnings component of 46% to 49% in 1996-2007.

While differences across countries need to be interpreted cautiously due to differences in earnings definition or data collection, lower levels of between-firm earnings dispersion in Sweden are often attributed to historically highly unionized labor market and the presence of collective wage bargaining agreements. In particular, after World War II the introduction of the so-called solidarity wage policy, which had as guiding principle "equal pay for equal work", significantly limited the capacity of firms to differentially pay their employees. However, several reforms over the last two decades have contributed to an increase in between-firm wage variation due to a more informal coordination in wage setting (see Skans, Edin, and Holmlund, 2009, and

[22]See their Figure 7.14. Skans, Edin, and Holmlund (2009) report the fraction of workers employed in plants with more than 25 employees in years $t-1$ and t who changed plant between $t-1$ and t.

[23]As a comparison, for Germany Fitzenberger and Garloff (2007) report yearly between-employers transition rates of 7.5% in the period 1976 to 1996 for male workers.

Akerman, Helpman, Itskhoki, Muendler, and Redding, 2013). It will be important to keep these features of the Swedish labor market in mind when interpreting the results.

Finally, comparing the first two columns (or the last two columns) of Table 1 shows that job movers are on average younger and more educated than workers who remain in the same firm. They also tend to work more in service sectors as opposed to manufacturing. In the last row we also see that firms with a non-zero fraction of job movers seem more productive, as their value added per worker is higher.

6 Empirical results (preliminary)

We now present results for static and dynamic models on the Swedish data. We start by describing the firm classes that we estimate using clustering.

6.1 Firm classes

As described in Section 4, we estimate firm classes using a weighted k-means algorithm to firms' log-earnings cdfs in 2002. To implement this estimator in practice we compute cdfs on a grid of 40 percentiles of the overall log-earnings distribution. We use the Hartigan-Wong algorithm, with 1000 random starting values.[24]

Table 2 provides summary statistics on the estimated firm classes, for $K = 10$. In the table we have ordered firm classes according to mean log-earnings in each class (although the ordering of the classes is arbitrary in our setting). Firm classes capture substantial heterogeneity between firms. The between-firm-class log-earnings variance is 0.0447, that is, 94% of the overall between-firm variance. This is important as this suggests that assuming homogeneity within each of the 10 classes does not result in major losses of information, at least in terms of variance of log-earnings.

There are also substantial differences between classes in terms of observables. While the lower classes (in terms of their mean log-earnings) show high percentages of high school dropouts and low percentages of workers with some college, the higher classes show the opposite pattern. We also observe that lower classes tend to have higher percentages of 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 earnings in these data. Workers above 51 years old are more evenly distributed between firm classes.

[24] We use the code "kmeansW" in the R package "FactoClass".

Table 2: Data description, by estimated firm classes

class:	1	2	3	4	5	6	7	8	9	10	all
number of workers	15,503	51,556	47,412	61,118	78,595	69,828	63,149	100,035	67,859	44,720	599,775
number of firms	5,474	6,425	5,144	3,891	4,506	3,895	2,838	3,745	4,013	3,895	43,826
number of firms $\geq$ 10	1,640	3,151	2,507	2,590	2,669	1,965	1,950	2,211	2,413	2,293	23,389
number of firms $\geq$ 50	107	445	290	655	532	339	501	495	517	457	4,338
mean firm reported size	12.2	22.88	20.73	49.22	37.21	35.97	66.83	55.77	50.42	58.22	37.59
median firm reported size	7	9	9	14	12	10	15	12	12	12	10
firm reported size for median worker	9	32	38	109	136	323	274	712	290	176	154
firm actual size for median worker	4	15	21	48	78	191	126	385	120	65	72
% high school drop out	28.2%	27.7%	28%	24.8%	26.3%	22.8%	20%	18.8%	10.3%	3.7%	20.6%
% high school graduates	61.6%	63.1%	64%	61.5%	62.6%	62.9%	57.4%	57.7%	44.5%	28%	56.7%
% some college	10.2%	9.2%	8.1%	13.7%	11%	14.3%	22.6%	23.5%	45.2%	68.3%	22.7%
% workers younger than 30	24.4%	19.8%	17.6%	20.8%	17.6%	15.8%	17.1%	13.3%	14.2%	15.4%	16.8%
% workers between 31 and 50	53.9%	54.2%	56.1%	55%	56.3%	57.4%	57.2%	58.1%	59.3%	62.7%	57.2%
% workers older than 51	21.8%	25.9%	26.3%	24.2%	26.1%	26.8%	25.6%	28.7%	26.5%	21.9%	26%
% workers in manufacturing	23.3%	40.2%	44.6%	46.4%	55.6%	54.3%	52.1%	49.8%	37.1%	19.9%	45.4%
% workers in services	39.8%	32.8%	28%	20.2%	16.6%	13.4%	13.5%	18.7%	36.7%	63.2%	25.3%
% workers in retail and trade	27%	19.3%	11.7%	30.3%	12.4%	8.9%	30.5%	8.2%	17.1%	15.8%	16.7%
% workers in construction	9.9%	7.6%	15.8%	3.1%	15.4%	23.3%	4%	23.4%	9.2%	1.1%	12.6%
mean log-earnings	9.67	9.91	10	10.04	10.09	10.17	10.18	10.25	10.4	10.65	10.18
variance of log-earnings	0.102	0.059	0.045	0.1	0.057	0.054	0.113	0.074	0.1	0.153	0.124
between-firm variance of log-earnings	0.0466	0.0043	0.0019	0.0038	0.0013	0.0013	0.0036	0.0017	0.004	0.0327	0.0475
mean log-value-added per worker	14.46	15.03	15.04	15.62	15.43	15.27	15.83	15.61	15.72	15.82	15.3

Notes: Males, fully employed in the same firm 2002 and 2004, continuously existing firms. Figures for 2002.

We observe that firm size tends to increase with firm class. In particular, firm sizes appear very small in class 1. This suggests that misclassification may be more likely in this case. There is also evidence of both between- and within-sector variation between classes, which is not monotonic in earnings levels. Lastly, log valued-added per worker tends to increase in firm class, suggesting that the ordering based on mean log wages broadly agrees with an ordering in terms of productivity. However, it is worth noting that the classes explain only 22.5% of the between-firm variance in log value-added per worker, suggesting that productivity differences within classes are substantial.

We next describe some patterns of mobility across firm classes. In Table E3 in Appendix E we report the number of movers between all pairs of classes. There is substantial worker mobility between firm classes. Moreover, there is evidence that log-earnings of job movers differ substantially, depending on the classes they come from and the classes they move to.

This is important, as our identification analysis above shows that heterogeneity in worker types across job movements is key for identification. In Figure E3 we report means of log-earnings for workers moving between class k and k' (x-axis) and for those moving from k' to k (y-axis), over 2002 and 2004 and for each pair of firm classes (k, k') with $k < k'$ (that is, on average k' firms have higher earnings than k firms). We see large differences in average earnings between upward and downward moves, with the latter being almost uniformly higher. Our results below will suggest that this pattern is mostly due to the strong sorting of high-earning workers in high-paying firms, which implies that workers moving downward (respectively, upward) are predominantly higher (resp., lower) types.

6.2 Results (UNDER CONSTRUCTION-PLEASE DO NOT CITE)

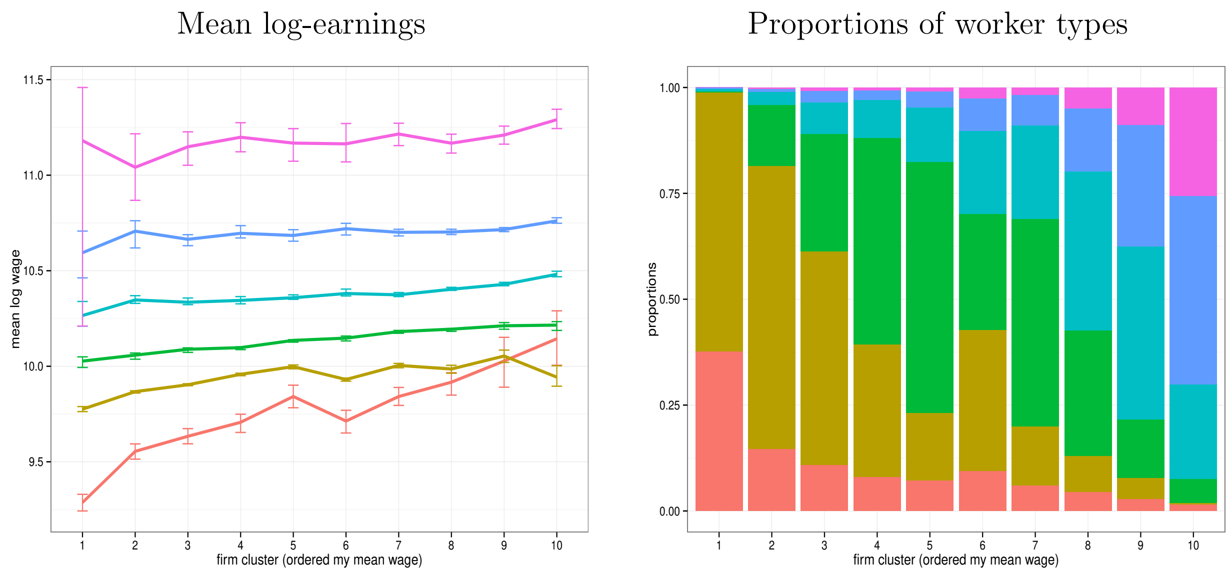
Static model. We start by estimating the static model on 2002 and 2004, based on a finite mixture specification for worker heterogeneity. We use $L = 6$ worker types. Given the estimated firm classes, we use the EM algorithm on the 2002-2004 job movers to estimate log-earnings distributions $\widehat{F}_{k\alpha}$. In our main specification log-earnings 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 use several strategies to pick sensible starting values.[25] Then, we estimate the worker type proportions $q_k(\alpha)$ from the 2002 data, using a second EM algorithm. The latter is numerically well-behaved as the likelihood is concave.

On the left panel of Figure 2 we report the estimates of the type-and-class-specific means of log-earnings. Different lines correspond to different worker types. On the x-axis, firm classes are ordered by mean log-earnings. The results show substantial heterogeneity of mean wages across workers within a firm class. At the same time, earnings variation across firm classes conditional on worker type seems more limited. An exception is the lowest worker type (in terms of average log-earnings), for whom earnings are substantially higher in firm class 10 compared to class 1. This finding points towards larger complementarity between low type workers and firm classes. Overall, however, the graph does not show strong evidence against an additive worker/firm specification.

On the right hand panel of Figure 2 we report the estimates of the proportions of worker types in each of the firm classes. The results show strong evidence that high worker types are matched to high firm classes, when both types and classes are ordered in terms of average

[25] We first take as starting value 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 2: Static model



Notes: The left graph plots mean log-earnings, by worker type and firm class. The $K = 10$ firm classes (on the x-axis) are ordered by mean log wage. The $L = 6$ worker types correspond to the 6 different colors. 95% confidence intervals based on the parametric bootstrap (200 replications). The right plot shows the proportions of worker types in each firm class.

log-earnings.

We next report the results of variance decompositions on simulated data and on the Swedish data. To simulate from the model we use the estimated parameters, and we condition on the firm identifiers and job movements observed in the data. The model allows simulating the wages of job movers in 2002 and 2004. In addition, we add serial correlation so as to approximately match the within-job log-earnings autocorrelation in the data. We use this simulated data to perform variance decomposition exercises as well as checks on the fit.

We estimate a linear regression of log-earnings on worker type indicators and firm class indicators on the simulated data. We then compute the components of the log-earnings variance explained by worker types (that is, the variance of the type-specific fixed-effects, which we denote as α) and firm classes (the variance of the class-specific fixed-effects ψ), and the contribution of the association between the two (twice the covariance of α and ψ). The first row in Table 3 shows that worker heterogeneity explains substantially more than firm heterogeneity according to our estimates. Indeed, the variance of the 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 on the right panel of

Figure 2.

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 a two-way fixed-effects type of regression on the Swedish data (referred to as "fixed-effects).[26] The results of the fixed-effects estimation are 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 in the literature using this type of regression.

There are several reasons why the results of these two specifications might differ. In order to better understand these differences, we repeat the variance decomposition exercises on a sample simulated from the static 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 in the Appendix. In contrast, the results obtained using fixed-effects show a strong negative correlation (-35%), in a data generating process where the true correlation is positive and substantial (46%). In addition, the fixed-effects results overestimate the contribution of firm heterogeneity to the variance of log-earnings substantially.

These results suggest the presence of a strong incidental bias in these data. With few job movers per firm, worker/firm correlations estimated using fixed effects may be downwardly biased, even negative when the true correlation is positive. This motivates the final exercise reported 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), the fixed-effects method recovers the correlation and the percentage of variance explained by the firms 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 fixed-effects method seems not to properly recover the amount of sorting, and more generally the contributions of worker and firm heterogeneity to

[26] We control for 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 the sample of movers between 2002 and 2004). Given the estimated firm fixed-effects, we then estimate individual fixed effects for the non-movers.

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			
This paper			0.7766	0.0473	0.1762	0.4598
Fixed-effects			0.9813	0.3014	-0.2826	-0.2599
			Simulated from the model			
This paper	1	1	0.7669	0.0466	0.1866	0.4934
Fixed-effects	1	1	1.0879	0.3447	-0.4326	-0.3532
			Simulated from the model without limited mobility			
Fixed-effects	4	1	0.8948	0.1602	-0.055	-0.0727
Fixed-effects	4	10	0.7816	0.053	0.1654	0.4064

Notes: Actual and simulated data. α is the worker (or type) fixed-effect, ψ is the firm (or class) fixed-effect. "This paper" is the static model with discrete worker heterogeneity. "min spell" is the minimum length of employment spells. "rep" is the number of job movers per firm, relative to the original dataset.

log-earnings dispersion, unless the data contain a large number of job movers per firm and long employment spells. Discrete heterogeneity seems to alleviate partly the incidental parameter problem in this data and can deal with short panels with relatively few job movements.

Dynamic Model. TO BE COMPLETED

7 Conclusion

In this paper we propose a framework designed for matched employer-employee datasets. Our aim is to build a bridge between reduced-form and structural approaches. We introduce empirical models which allow for interaction effects and dynamics, hence capturing mechanisms that have been emphasized in theoretical work. We have characterized conditions for identification in these models, and developed estimators for finite mixtures and interactive regression models.

Our two-step estimation approach could be useful in structural settings, where joint estimation of the distribution of two-sided heterogeneity and the structural parameters may be computationally prohibitive. In companion work (Bonhomme, Lamadon, and Manresa, 2015), we are currently further studying the theoretical properties of such approaches based on an initial clustering step.

Finally, we expect our methods will be useful in studies of workers and firms, where dimension reduction could help studying particular occupations or short periods of time for example, but also in other contexts involving matched data such as in economics of education, urban economics, or finance.

References

Abadie, A., M. M. Chingos, and M. R. West (2014): "Endogenous Stratification in Randomized Experiments," *Working Paper.*

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 (2015): "Sorting Between and Within Industries: A Testable Model of Assortative Matching," Discussion paper, Labor Dynamics Institute, Cornell University.

Abowd, J. M., I. M. Schmutte, and K. L. McKinney (2015): "Modeling Endogenous Mobility in Wage Determination," *Working paper.*

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.

Akerman, A., E. Helpman, O. Itskhoki, M.-A. Muendler, and S. Redding (2013): "Sources of Wage Inequality," *American Economic Review, Papers and Proceedings.*, 103(3), 214–219.

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.*

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2014): "It's Where You Work: Increases in Earnings Dispersion Across Establishments and Individuals in the U.S.," *Working Paper.*

Baumgarten, D., G. Felbermayr, and S. Lehwald (2014): "Drivers of Wage Inequality in Germany: Trade, Technology, or Institutions?," *Working Paper.*

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, pp. 813–846.

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 (2016): "Nonparametric estimation of finite mixtures from repeated measurements," *Journal of the Royal Statistical Society, Series B, forthcoming.*

Bonhomme, S., T. Lamadon, and E. Manresa (2015): "Approximate Clustering," *Working Paper.*

Bonhomme, S., and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83(3), 1147–1184.

Burdett, K., and M. G. Coles (2003): "Wage/tenure contracts and equilibrium search," *Econometrica*, 71(5), 1377–1404.

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

Canay, I., A. Santos, and A. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models with Endogeneity," *Econometrica*, 81(6), 97–111.

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.*

Delacroix, A., and S. Shi (2006): "Directed Search on the Job and the Wage Ladder," *Int. Econ. Rev.*, 47(2), 651–699.

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.

Dey, M. S., and C. J. Flinn (2005): “An Equilibrium Model of Health Insurance Provision and Wage Determination,” *Econometrica*, 73(2), 571–627.

Eeckhout, J., and P. Kircher (2011): “Identifying sorting in theory,” *Review of Economic Studies*, 78(3), 872–906.

Engbom, N., and C. Moser (2015): “Earnings Inequality and the Minimum Wage: Evidence from Brazil,” Discussion paper.

Finkelstein, A., M. Gentzkow, and H. Williams (2014): “Sources of Geographic Variation in Health Care: Evidence from Patient Migration,” *Working Paper.*

Fitzenberger, B., and A. Garloff (2007): “Descriptive Evidence on Labor Market Transitions and the Wage Structure in Germany,” *Int. Econ. Rev.*, 227(2), 115–152.

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

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

Goldschmidt, D., and J. Schmieder (2015): “The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure,” *Working Paper.*

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.

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *J. Econom.*, 171(1), 32–44.

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

Kasahara, H., and K. Shimotsu (2014): "Nonparametric identification and estimation of the number of components in multivariate mixtures," *Journal of the Royal Statistical Society, Series B*, 76, 97–111.

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," Discussion paper.

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*.

Lise, J., and J.-M. Robin (2013): "The macro-dynamics of sorting between workers and firms," *Working Paper*.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

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.

Moscarini, G., and F. Postel-Vinay (2009): "Non Stationary Search Equilibrium," *Working Paper*.

Postel-Vinay, F., and J.-m. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

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

Shephard, A. (2011): "Equilibrium search and tax credit reform," Discussion paper.

Shi, S. (2008): "Directed Search for Equilibrium Wage-Tenure Contracts," *Econometrica*, 77(2), 561–584.

Shimer, R. (2001): "The assignment of workers to jobs in an economy with coordination frictions," *NBER Work. Pap. Ser.*

——— (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.

Skans, O. N., P.-A. Edin, and B. Holmlund (2009): "Wage Dispersion Between and Within Plants: Sweden 1985-2000," in *The Structure of Wages: An International Comparison*, ed. by Edward P. Lazear and Kathryn L. Shaw, pp. 217–260. University of Chicago Press.

Song, J., D. J. Price, F. Guvenen, and N. Bloom (2015): "Firming up Inequality," *Working Paper*.

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

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 $k \in \{1, ..., K\}$, and let $(k_1, ..., k_R)$, $(\widetilde{k}_1, ..., \widetilde{k}_R)$ as in Assumption 3, with $k_1 = k$. From (2) we have, considering workers who move between k_r and $\widetilde{k}_{r'}$ for some $r \in \{1, ..., R\}$ and $r' \in \{r-1, r\}$:

$$\Pr\left[Y_{i1} \leq y_1, Y_{i2} \leq y_2 \,|\, k_{i1} = k_r, k_{i2} = \widetilde{k}_{r'}, m_{i1} = 1\right] = \sum_{\alpha=1}^{L} p_{k_r,\widetilde{k}_{r'}}(\alpha) F_{k_r,\alpha}(y_1) F^m_{\widetilde{k}_{r'},\alpha}(y_2). \tag{A1}$$

Consider a set of M values for y_1, and the same set of values for y_2, that satisfy Assumption 3 *ii*). Writing (A1) in matrix notation we obtain:

$$A(k_r, \widetilde{k}_{r'}) = F(k_r) D(k_r, \widetilde{k}_{r'}) F^m(\widetilde{k}_{r'})^\intercal, \tag{A2}$$

where $A(k_r, \widetilde{k}_{r'})$ is $M \times M$ with generic element:

$$\Pr\left[Y_{i1} \leq y_1, Y_{i2} \leq y_2 \,|\, k_{i1} = k_r, k_{i2} = \widetilde{k}_{r'}, m_{i1} = 1\right],$$

$F(k_r)$ is $M \times K$ with element $F_{k_r,\alpha}(y_1)$, $F^m(\widetilde{k}_{r'})$ is $M \times K$ with element $F^m_{\widetilde{k}_{r'},\alpha}(y_2)$, and $D(k_r, \widetilde{k}_{r'})$ is $K \times K$ diagonal with element $p_{k_r,\widetilde{k}_{r'}}(\alpha)$.

Note that $A(k_r, \widetilde{k}_{r'})$ has rank K by Assumption 3 *ii*). Consider a singular value decomposition of $A(k_1, \widetilde{k}_1)$:

$$A(k_1, \widetilde{k}_1) = F(k_1) D(k_1, \widetilde{k}_1) F^m(\widetilde{k}_1)^\intercal = USV^\intercal,$$

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

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

$B(k_r, \widetilde{k}_{r'})$ and $Q(k_r)$ are non-singular by Assumption 3 *ii*). Moreover, we have, for all $r \in \{1, ..., R\}$:

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

Let $E_r = B(k_r, \widetilde{k}_r) B(k_{r+1}, \widetilde{k}_r)^{-1}$. We thus have:

$$E_1 E_2 ... E_R = Q(k_1) D(k_1, \widetilde{k}_1) D(k_2, \widetilde{k}_1)^{-1} ... D(k_R, \widetilde{k}_R) D(k_1, \widetilde{k}_R)^{-1} Q(k_1)^{-1}.$$

The eigenvalues of this matrix are all distinct by Assumption 3 i), so $Q(k_1)=Q(k)$ is identified up to multiplication by a diagonal matrix and permutation of its columns.

Now, note that $F(k)=UU^{\intercal}F(k)$, so:

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

is identified up to multiplication by a diagonal matrix and permutation of its columns. Hence $F_{k\alpha}(y_1)\lambda_\alpha$ is identified up to a choice of labelling, where $\lambda_\alpha\neq 0$ is a scale factor. Taking $y_1=+\infty$, λ_α is identified, so $F_{k\alpha}(y_1)$ is identified up to labelling. As a result, $F_{k,\sigma(\alpha)}(y_1)$ is identified for some permutation $\sigma:\{1,...,L\}\to\{1,...,L\}$. To identify $F_{k,\sigma(\alpha)}$ at a point y different from the grid of M values considered so far, simply augment the set of values with y as an additional (i.e., $(M+1)$th) value, and apply the above arguments.

Let now $k'\neq k$, and let $(k_1,...,k_R)$, $(\widetilde{k}_1,...,\widetilde{k}_R)$, be an alternating cycle such that $k_1=k$ and $k'=k_r$ for some r, by Assumption 3 i). We have:

$$A(k,\widetilde{k}_1)=F(k)D(k,\widetilde{k}_1)F^m(\widetilde{k}_1)^{\intercal}.$$

As $F_{k,\sigma(\alpha)}$ is identified and $F(k)$ has rank K:

$$p_{k,\widetilde{k}_1}(\sigma(\alpha))F^m_{\widetilde{k}_1,\sigma(\alpha)}(y_2)$$

is identified, so by taking $y_2=+\infty$, both $p_{k,\widetilde{k}_1}(\sigma(\alpha))$ and $F^m_{\widetilde{k}_1,\sigma(\alpha)}$ are identified. Next we have:

$$A(k_2,\widetilde{k}_1)=F(k_2)D(k_2,\widetilde{k}_1)F^m(\widetilde{k}_1)^{\intercal},$$

so, using similar arguments, $p_{k_2,\widetilde{k}_1}(\sigma(\alpha))$ and $F_{\widetilde{k}_2,\sigma(\alpha)}$ are identified. By induction, $p_{k_r,\widetilde{k}_{r'}}(\sigma(\alpha))$, $F_{k_r,\sigma(\alpha)}$, and $F^m_{\widetilde{k}_{r'},\sigma(\alpha)}$ are identified for all r and $r'\in\{r-1,r\}$. As $k'=k_r$, it follows that $F_{k',\sigma(\alpha)}$ and $F^m_{k',\sigma(\alpha)}$ are identified. Note that the same argument implies that $F^m_{k,\sigma(\alpha)}$ is identified.

Lastly, let $(k,k')\in\{1,...,K\}^2$. Then, from:

$$A(k,k')=F(k)D(k,k')F^m(k')^{\intercal},$$

and from the fact that $F_{k,\sigma(\alpha)}$ and $F^m_{k',\sigma(\alpha)}$ are both identified, and that $F(k)$ and $F^m(k')$ have rank K by Assumption 3 ii), it follows that $p_{kk'}(\sigma(\alpha))$ is identified.

Proof of Corollary 1. By Theorem 1 there exists a permutation $\sigma:\{1,...,L\}\to\{1,...,L\}$ such that $F_{k,\sigma(\alpha)}$ is identified for all k,α. Now we have, writing (3) for the K worker types and the M values of y_1 given by Assumption 3 ii) in matrix form:

$$H(k)=F(k)P(k),$$

where $H(k)$ has generic element $\Pr[Y_{i1} \leq y_1 \,|\, k_{i1} = k]$, the $L \times 1$ vector $P(k)$ has generic element $q_k(\sigma(\alpha))$, and the columns of $F(k)$ have been ordered with respect to σ. By Assumption 3 *ii*), $F(k)$ has rank K, from which it follows that:

$$P(k) = [F(k)^{\intercal} F(k)]^{-1} F(k)^{\intercal} H(k)$$

is identified. So $q_k(\sigma(\alpha))$ is identified.

Proof of Corollary 2. We start by listing the required assumptions.

Definition A1. *An augmented alternating cycle of length R is a pair of sequences of firm classes and log-earnings values $(k_1, y_1, ..., k_R, y_R)$ and $(\widetilde{k}_1, \widetilde{y}_1, ..., \widetilde{k}_R, \widetilde{y}_R)$, with $k_{R+1} = k_1$ and $y_{R+1} = y_1$, such that $p_{y_r,\widetilde{y}_r,k_r,\widetilde{k}_r}(\alpha) \neq 0$ and $p_{y_{r+1},\widetilde{y}_r,k_{r+1},\widetilde{k}_r}(\alpha) \neq 0$ for all r in $\{1, ..., R\}$ and α in $\{1, ..., L\}$.*

Assumption A1. *(mixture model, dynamic)*
i) For any two firm classes $k \neq k'$ in $\{1, ..., K\}$ and any two log-earnings values $y \neq y'$, there exists an augmented alternating cycle $(k_1, y_1, ..., k_R, y_R)$ and $(\widetilde{k}_1, \widetilde{y}_1, ..., \widetilde{k}_R, \widetilde{y}_R)$, such that $(k_1, y_1) = (k, y)$, and $(k_r, y_r) = (k', y')$ for some r, and such that the scalars $a(1), ..., a(L)$ are all distinct, where:

$$a(\alpha) = \frac{p_{y_1,\widetilde{y}_1,k_1,\widetilde{k}_1}(\alpha) p_{y_2,\widetilde{y}_2,k_2,\widetilde{k}_2}(\alpha) ... p_{y_R,\widetilde{y}_R,k_R,\widetilde{k}_R}(\alpha)}{p_{y_2,\widetilde{y}_1,k_2,\widetilde{k}_1}(\alpha) p_{y_3,\widetilde{y}_2,k_3,\widetilde{k}_2}(\alpha) ... p_{y_1,\widetilde{y}_R,k_1,\widetilde{k}_R}(\alpha)}.$$

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

ii) For a suitable finite set of values for y_1 and y_4, which includes $(+\infty, +\infty)$, and for all r in $\{1, ..., R\}$, the matrices $A(y_r, \widetilde{y}_r, k_r, \widetilde{k}_r)$ and $A(y_r, \widetilde{y}_{r+1}, k_r, \widetilde{k}_{r+1})$ have rank K, where:

$$A(y, y', k, k') = \left\{\Pr\left[Y_{i1} \leq y_1, Y_{i4} \leq y_4 \,|\, Y_{i2} = y, Y_{i3} = y', k_{i2} = k, k_{i3} = k', m_{i2} = 1\right]\right\}_{(y_1,y_4)}.$$

We are now in position to prove Corollary 2.

Part (i) is a direct application of Theorem 1, under Assumption A1.

For part (ii) we have, from (6):

$$\Pr[Y_{i1} \leq y_1 \,|\, Y_{i2} = y_2, k_{i1} = k_{i2} = k, m_{i1} = 0] = \sum_{\alpha=1}^{L} G^f_{y_2,k\alpha}(y_1) \pi_{y_2,k}(\alpha),$$

where:

$$\pi_{y_2,k}(\alpha) = \frac{q_k(\alpha) f_{k\alpha}(y_2)}{\sum_{\widetilde{\alpha}=1}^{L} q_k(\widetilde{\alpha}) f_{k\widetilde{\alpha}}(y_2)}$$

are the posterior probabilities of worker types given $Y_{i2} = y_2$, $k_{i2} = k$, and $m_{i1} = 0$, with $f_{k\alpha}$ denoting the distribution function of log-earnings given $\alpha_i = \alpha$, $k_{i2} = k$, and $m_{i1} = 0$, and $q_k(\alpha)$ denoting the proportion of workers of type α with $k_{i2} = k$ and $m_{i1} = 0$.

Given the rank condition on the $M \times K$ matrix with generic element $G^f_{y_2,k\alpha}(y_1)$, which is identified up to labelling of α, $\pi_{y_2,k}(\alpha)$ are thus identified up to the same labelling. Hence:

$$q_k(\alpha) = \Pr\left[\alpha_i = \alpha \,|\, k_{i2} = k, m_{i1} = 0\right] = \mathbb{E}\left[\pi_{Y_{i2},k}(\alpha) \,|\, k_{i2} = k, m_{i1} = 0\right]$$

is also identified up to labelling. By Bayes' rule, the second period's log-earnings cdf:

$$F_{k\alpha}(y_2) = \Pr\left[Y_{i2} \leq y_2 \,|\, \alpha_i = \alpha, k_{i2} = k, m_{i1} = 0\right] = \mathbb{E}\left[\frac{\pi_{Y_{i2},k}(\alpha)}{q_k(\alpha)}\mathbf{1}\{Y_{i2} \leq y_2\} \,\middle|\, k_{i2} = k, m_{i1} = 0\right]$$

is thus also identified up to labelling. Similarly, the log-earnings cdfs in all other periods can be uniquely recovered up to labelling, 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 identified as:

$$\Pr\left[k_{i3} = k' \,|\, \alpha_i = \alpha, Y_{i2} = y_2, k_{i2} = k, m_{i2} = 1\right] = \frac{\int p_{y_2y_3,kk'}(\alpha) q_{kk'}(y_2, y_3) dy_3}{\sum_{\widetilde{k}=1}^{K} \int p_{y_2y_3,k\widetilde{k}}(\alpha) q_{k\widetilde{k}}(y_2, y_3) dy_3},$$

where $q_{kk'}(y_2, y_3)$ is defined by:

$$\int_{-\infty}^{y} q_{kk'}(y_2, y_3) dy_3 = \Pr\left[Y_{i3} \leq y, k_{i3} = k' \,|\, Y_{i2} = y_2, k_{i2} = k, m_{i2} = 1\right].$$

Finally, note that $q_k(\alpha)$ and $f_{k\alpha}$ are conditional on the worker not moving between periods 1 and 2 (i.e., $m_{i1} = 0$). One could recover unconditional probabilities by also using job movers in the first periods ($m_{i1} = 1$), although we do not provide details here.

Proof of Proposition 1. Note that (13) is equivalent to the following k-means problem:

$$\min_{k(1),...,k(J),H_1,...,H_K} \sum_{i=1}^{N} \int \left(\mathbf{1}\{Y_{i1} \leq y_1\} - H_{k(j_{i1})}(y_1)\right)^2 d\mu(y_1).$$

We now verify Assumptions 1 and 2 in Bonhomme and Manresa (2015). Note that their setup allows for unbalanced structures (that is, different n_j across j) provided the assumptions are formulated in terms of the minimum firm size in the sample: $n = \min_j n_j$. Assumptions 1a and 1c are satisfied because $\mathbf{1}\{Y_{i1} \leq y_1\}$ is bounded. Assumptions 1d, 1e, and 1f hold because of Assumption 4 (i). Assumptions 2a and 2b hold by Assumptions 4 (ii) and (iii). Finally, Assumptions 2c and 2d are also satisfied by Assumption 4 (i) and boundedness of $\mathbf{1}\{Y_{i1} \leq y_1\}$. Theorems 1 and 2 in Bonhomme and Manresa (2015) and Assumption 4 (iv) imply the result.

B Complements

B.1 Identification of log-earnings distributions: an example

Here we consider a setting where worker types and firm classes are ordered (e.g., by their productivity) and their is strong positive assortative matching in the economy. Formally, we suppose that $K = L$, that $q_k(\alpha) \neq 0$ if and only if $|k-\alpha| \leq 1$, and that $p_{kk'}(\alpha) \neq 0$ if and only if $(|k-\alpha| \leq 1, |k'-\alpha| \leq 1)$. Borrowing the notation from the proof of Theorem 1, assume that all matrices $F(k)$ and $F^m(k')$ have full-column rank K, for all k, k'.

Then $\operatorname{rank} A(1,2) = \operatorname{rank} A(2,1) = 2$. It follows as in the proof of Theorem 1 that (F_{11}, F_{12}), (F_{21}, F_{22}), (F_{11}^m, F_{12}^m), and (F_{21}^m, F_{22}^m), are identified up to a choice of labelling.

Likewise, $\operatorname{rank} A(2,3) = \operatorname{rank} A(3,2) = 3$. It follows that, for some $(\alpha_1, \alpha_2, \alpha_3)$, $(F_{2\alpha_1}, F_{2\alpha_2}, F_{2\alpha_3})$ and $(F_{3\alpha_1}, F_{3\alpha_3}, F_{3\alpha_3})$ are identified, and similarly for the corresponding F^m's.

As $F(2)$ has full column rank, one can pin down which one of the types $(\alpha_1, \alpha_2, \alpha_3)$ are equal to 1 or 2. Without loss of generality, let $\alpha_1 = 1$ and $\alpha_2 = 2$. Set $\alpha_3 = 3$. Then (F_{21}, F_{22}, F_{23}) and (F_{31}, F_{32}, F_{33}) are identified, and similarly for the corresponding F^m's.

Continuing the argument we identify: (F_{11}, F_{12}), (F_{11}^m, F_{12}^m), (F_{12}, F_{22}, F_{32}), $(F_{12}^m, F_{22}^m, F_{32}^m)$, and so on, until $(F_{K-1,K-2}, F_{K-1,K-1}, F_{K-1,K})$, $(F_{K-1,K-2}^m, F_{K-1,K-1}^m, F_{K-1,K}^m)$, $(F_{K,K-1}, F_{KK})$, and finally $(F_{K,K-1}^m, F_{KK}^m)$.

The other $F_{k\alpha}$'s are not identified. These correspond to the (k,α) combinations such that $q_k(\alpha) = 0$. In this example, without additional structure one cannot assess the earnings effects of randomly allocating workers to jobs, for instance.

B.2 Nonparametric identification for continuous worker types

The analysis here is closely related to Hu and Schennach (2008). Let us define the following operators:

$$\begin{aligned}
L_{kk'}g(y_1) &= \int f_{kk'}(y_1, y_2) g(y_2) dy_2, \\
A_k h(y_1) &= \int f_{k\alpha}(y_1) h(\alpha) d\alpha, \\
B_{k'}^m g(\alpha) &= \int f_{k'\alpha}^m(y_2) g(y_2) dy_2, \\
D_{kk'}h(\alpha) &= p_{kk'}(\alpha) h(\alpha).
\end{aligned}$$

In operator form, (11) becomes:

$$L_{kk'} = A_k D_{kk'} B_{k'}^m. \tag{B3}$$

Consider an alternating cycle of length $R = 2$. Suppose that A_k are $B_{k'}^m$ are *injective*, and that $p_{kk'}(\alpha) > 0$, for all $(k,k') \in \{k_1, k_2\} \times \{\widetilde{k}_1, \widetilde{k}_2\}$. Lastly, suppose for all $\alpha \neq \alpha'$:

$$\frac{p_{k_1\widetilde{k}_1}(\alpha) p_{k_2\widetilde{k}_2}(\alpha)}{p_{k_1\widetilde{k}_2}(\alpha) p_{k_2\widetilde{k}_1}(\alpha)} \neq \frac{p_{k_1\widetilde{k}_1}(\alpha') p_{k_2\widetilde{k}_2}(\alpha')}{p_{k_1\widetilde{k}_2}(\alpha') p_{k_2\widetilde{k}_1}(\alpha')}. \tag{B4}$$

A condition similar to (B4) arises in the analysis of Hu and Shum (2012). Operator injectivity is related to completeness in the literature on nonparametric instrumental variables estimation. It is a nonparametric analogue of a rank condition. However, injectivity or completeness may be difficult to test formally, and they are high-level conditions (Canay, Santos, and Shaikh, 2013). With $T=2$, injectivity requires α_i to be one-dimensional.

Under these assumptions the operators $L_{kk'}$, A_k, $B^m_{k'}$, and $D_{kk'}$ are invertible. Moreover, analogously to Hu and Schennach (2008) one can show that the following spectral decomposition is unique:

$$L_{k_1\widetilde{k}_1}L^{-1}_{k_2\widetilde{k}_1}L_{k_2\widetilde{k}_2}L^{-1}_{k_1\widetilde{k}_2}=A_{k_1}\left[D_{k_1\widetilde{k}_1}D^{-1}_{k_2\widetilde{k}_1}D_{k_2\widetilde{k}_2}D^{-1}_{k_1\widetilde{k}_2}\right]A^{-1}_{k_1}.$$

This implies that the density $f_{k\alpha}(y_1)$ is identified up to an arbitrary one-to-one transformation of α. A possible scaling is obtained if there is a known functional $\mathcal{F}$ such that $\mathcal{F}f_{k\alpha}$ is monotone in α. The functional $\mathcal{F}$ may depend on k. As an example, identification is achieved if $\mathbb{E}\left[Y_{i1}\,|\,\alpha_i=\alpha,k_{i1}=k\right]$ is monotone in α. In that case one may normalize $\alpha_i=\mathbb{E}\left[Y_{i1}\,|\,\alpha_i=\alpha,k_{i1}=k\right]$.

B.3 Separation condition in bi-clustering

We consider the static mixture model with discrete worker types. Let $G_{kk'}$ denote the bivariate cdf on the left-hand side of (14). In (14) the separation condition is the following: for all k,k' there exists k'' such that $G_{kk''}\neq G_{k'k''}$ or $G_{k''k}\neq G_{k''k'}$. The following result shows that this separation condition is weaker than the one in the cross-sectional case, see Assumption 4 (*iii*).

Corollary B1. *Let $k\neq k'$ such that, for all k'', $G_{kk''}=G_{k'k''}$ and $G_{k''k}=G_{k''k'}$. Suppose that $F(k'')$ and $F^m(k'')$ have rank K for all k'', and that there exists k_1,k_2 such that $p_{k,k_2}(\alpha)>0$ and $p_{k_1,k}(\alpha)>0$ for all α. Then, $F_{k\alpha}=F_{k'\alpha}$ and $F^m_{k\alpha}=F^m_{k'\alpha}$ for all $\alpha\in\{1,...,L\}$.*

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

$$F(k)D(k,k'')F^m(k'')^\intercal \quad = \quad F(k')D(k',k'')F^m(k'')^\intercal.$$

Take $k''=k_2$. By assumption, $F^m(k_2)$ has rank K. So:

$$F(k)D(k,k_2)=F(k')D(k',k_2).$$

We thus get $F_{k\alpha}(y_1)p_{k,k_2}(\alpha)=F_{k'\alpha}(y_1)p_{k',k_2}(\alpha)$, so taking $y_1=+\infty$ we have $p_{k',k_2}(\alpha)=p_{k,k_2}(\alpha)>0$ and $F_{k\alpha}=F_{k'\alpha}$, for all $\alpha\in\{1,...,L\}$. Similarly, from $G_{k''k}=G_{k''k'}$ and the assumption that $F(k_1)$ has rank K we obtain that $F^m_{k\alpha}=F^m_{k'\alpha}$. ∎

Corollary B1 shows that, provided that type-specific earnings distributions differ, information from the earnings sequences of job movers can allow identifying firm classes even when the cross-sectional earnings information is insufficient. Note that the assumptions in the corollary are weaker than those in Theorem 1.

B.4 Time-varying firm classes

To outline how to estimate time-dependent firm classes $k_t(j)$, note that the classes in period 1 can be consistently estimated using (13). In the second period, one can estimate the period-specific classification by solving the following k-means problem:

$$\min_{k_2(1),\ldots,k_2(J),H_{11},\ldots,H_{KK}} \sum_{j=1}^{J} n_j \int \left(\widehat{F}_{2j}(y) - H_{\widehat{k}_1(j),k_2(j)}(y)\right)^2 d\mu(y), \tag{B5}$$

where $\widehat{F}_{2j}$ denotes the log-earnings cdf in period 2, and $\widehat{k}_1(j)$ are estimates from (13). This may be iterated until the last period of the panel.

B.5 Estimation on T periods

Here we outline estimation in models with T periods. The static model being a special case of the dynamic one, we focus on the latter.

Consider the dynamic model with a finite mixture specification for worker types. The estimation first step is as in (13). In practice one may sum the objective function over the T periods. With the class estimates $\widehat{k}_{it}$ at hand, in the second step we estimate the mixture model using maximum likelihood. The pieces of the likelihood function are as follows, where for simplicity we assume that observed characteristics X_{it} are strictly exogenous. Also, we explicitly indicate t as a conditioning variable, to emphasize that all distributions may depend on calendar time.

- Initial condition, types: $\Pr[\alpha_i = \alpha \,|\, k_{i1}, X_{i1}; \theta_1]$.
- Initial condition, log-earnings: $\Pr[Y_{i1} \leq y_1 \,|\, \alpha_i, k_{i1}, X_{i1}; \theta_2]$.
- Transitions, mobility: $\Pr[m_{it} = m \,|\, Y_{it}, \alpha_i, k_{it}, X_{it}, t; \theta_3]$.
- Transitions, classes: $\Pr[k_{i,t+1} = k' \,|\, Y_{it}, \alpha_i, k_{it}, X_{it}, m_{it} = 1, t; \theta_4]$.
- Transitions, log-earnings: $\Pr[Y_{i,t+1} \leq y_{t+1} \,|\, Y_{it}, \alpha_i, k_{i,t+1}, k_{it}, X_{i,t+1}, m_{it} = m, t; \theta_5]$.

C Interactive regression models

C.1 Models and identification

Static model. Consider the nonstationary static model (1) on $T = 2$ periods. Note that multiplying (1) by $\tau_t(k_{it}) = 1/b_t(k_{it})$, taking means for job movers, and taking time differences yields:

$$\mathbb{E}\left[Z_i\left(\tau_2(k_{i2})Y_{i2} - \tau_1(k_{i1})Y_{i1} - \widetilde{a}_2(k_{i2}) + \widetilde{a}_1(k_{i1}) - X_{i2}'\widetilde{c}_2(k_{i2}) + X_{i1}'\widetilde{c}_1(k_{i1})\right) \,|\, m_{i1} = 1\right] = 0, \tag{C6}$$

where $\widetilde{a}_t(k) = \tau_t(k)a_t(k)$, and $\widetilde{c}_t(k) = \tau_t(k)c_t$. The vector Z_i stacks together X_{i1}, X_{i2}, as well as all k_{i1} and k_{i2} dummies and their interactions, the interactions between X_{i1} and k_{i1} dummies, and those between X_{i2} and the k_{i2} dummies.[27]

Note that (C6) is linear in parameters. Linearity is an important property in order to develop a practical estimator. Let us fix, without loss of generality, $a_1(1) = 0$ and $b_1(1) = 1$. Our estimator in the next subsection will be invariant to the choice of normalization. Let A be the $(2d_xK + K^2) \times (2d_xK + 4K - 2)$ matrix that corresponds to the linear system in (C6), with d_x denoting the dimension of X_{it}. The order condition for identification in (C6) requires $K \geq 4$. We have the following result.

Theorem C1. *Consider model (1) with $T = 2$ and $\mathbb{E}(\varepsilon_{it} \,|\, \alpha_i, k_{i1}, k_{i2}, X_i, m_{i1}) = 0$, where $X_i = (X_{i1}, X_{i2})$. Suppose that $b_t(k) \neq 0$ for all t, k.*

(i) If A has maximal rank then the $b_t(k)$, $a_t(k)$, and c_t are all identified. Moreover, the means $\mathbb{E}(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1)$ and $\mathbb{E}(\alpha_i \,|\, k_{i1} = k_{i2} = k, m_{i1} = 0)$ are identified.

(ii) If, in addition to (i), $\mathrm{Cov}(\varepsilon_{i1}, \varepsilon_{i2} \,|\, k_{i1}, k_{i2}, m_{i1} = 1) = 0$, then $\mathrm{Var}(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1)$ are identified.

(ii) If, in addition to (i) and (ii), $\mathbb{E}\left(\varepsilon_{i1}^2 \,|\, k_{i1}, k_{i2}, m_{i1}\right) = \mathbb{E}\left(\varepsilon_{i1}^2 \,|\, k_{i1}\right)$ and $\mathbb{E}\left(\varepsilon_{i2}^2 \,|\, k_{i1}, k_{i2}, m_{i1}\right) = \mathbb{E}\left(\varepsilon_{i2}^2 \,|\, k_{i2}\right)$, then $\mathrm{Var}(\alpha_i \,|\, k_{i1} = k_{i2} = k, m_{i1} = 0)$, $\mathrm{Var}(\varepsilon_{i1} \,|\, k_{i1} = k)$, and $\mathrm{Var}(\varepsilon_{i2} \,|\, k_{i2} = k)$ are identified.

Proof. Part (i). If A has maximal rank then (C6) identifies the $\tau_t(k)$, $\widetilde{a}_t(k)$ and $\widetilde{c}_t(k)$. Hence the $b_t(k)$, $a_t(k)$, and c_t, are identified.[28] Identification of the means of α_i conditional on $m_{i1} = 0$ or $m_{i1} = 1$ then follows directly. For example, we have:

$$\mathbb{E}(\alpha_i \,|\, k_{i1}, k_{i2}, m_{i1}) = \mathbb{E}\left(\tau_1(k_{i1})Y_{i1} - \widetilde{a}_1(k_{i1}) - X_{i1}'\widetilde{c}_1(k_{i1}) \,|\, k_{i1}, k_{i2}, m_{i1}\right).$$

Part (ii). Let $\widetilde{Y}_{it} = Y_{it} - X_{it}'c_t$. If $\mathrm{Cov}(\varepsilon_{i1}, \varepsilon_{i2} \,|\, k_{i1}, k_{i2}, m_{i1} = 1) = 0$ then:

$$\mathrm{Var}\left(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1\right) = \tau_1(k)\tau_2(k')\,\mathrm{Cov}\left(\widetilde{Y}_{i1}, \widetilde{Y}_{i2} \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1\right)$$

is identified.

Part (iii). If $\mathbb{E}\left(\varepsilon_{i1}^2 \,|\, k_{i1}, k_{i2}, m_{i1}\right) = \mathbb{E}\left(\varepsilon_{i1}^2 \,|\, k_{i1}\right)$ then:

$$\mathrm{Var}(\varepsilon_{i1} \,|\, k_{i1} = k) = \mathrm{Var}\left(\widetilde{Y}_{i1} \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1\right) - b_1(k)^2\,\mathrm{Var}\left(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1\right)$$

is identified, and likewise for $\mathrm{Var}(\varepsilon_{i2} \,|\, k_{i2} = k)$. Lastly:

$$\mathrm{Var}(\alpha_i \,|\, k_{i1} = k_{i2} = k, m_{i1} = 0) = \tau_1^2(k)\left[\mathrm{Var}\left(\widetilde{Y}_{i1} \,|\, k_{i1} = k_{i2} = k, m_{i1} = 0\right) - \mathrm{Var}(\varepsilon_{i1} \,|\, k_{i1} = k)\right]$$

is thus identified.

∎

[27] An even richer set of instruments would also include interactions between X's and interactions of k_{i1} and k_{i2} dummies.

[28] By the same token one could also identify firm-class-specific coefficients $c_t(k)$.

Dynamic model. An interactive dynamic model on four periods is as follows, where we abstract from covariates for simplicity. We write:

$$\begin{aligned} Y_{it} &= a_t^s(k) + b_t(k)\alpha_i + \varepsilon_{it}, \quad t = 1, ..., 4, \\ &\text{if} \quad m_{i1} = 0,\, m_{i2} = 0,\, m_{i3} = 0, \end{aligned} \tag{C7}$$

for workers who remain in the same firm of class k in all periods, where "s" stands for "stayers".

Next, we consider workers who remain in the same firm of class k in periods 1 and 2 and move to a firm of class k' in periods 3 and 4. We specify their log-earnings as follows:

$$\begin{aligned} Y_{i1} &= a_1^s(k) + \rho_{1|2}(a_2^m(k) - a_2^s(k)) + \rho_{1|2}\xi_2(k') + b_1(k)\alpha_i + \varepsilon_{i1}, \\ Y_{i2} &= a_2^m(k) + \xi_2(k') + b_2(k)\alpha_i + \varepsilon_{i2}, \\ Y_{i3} &= a_3^m(k') + \xi_3(k) + b_3(k')\alpha_i + \varepsilon_{i3}, \\ Y_{i4} &= a_4^s(k') + \rho_{4|3}(a_3^m(k') - a_3^s(k')) + \rho_{4|3}\xi_3(k) + b_4(k')\alpha_i + \varepsilon_{i4}, \\ &\text{if} \quad m_{i1} = 0,\, m_{i2} = 1,\, m_{i3} = 0, \end{aligned} \tag{C8}$$

where "m" stands for "movers". In (C7) and (C8) we assume that:

$$\mathbb{E}\left(\varepsilon_{it} \,|\, \alpha_i, k_{i1}, k_{i2}, k_{i3}, k_{i4}, m_{i1}, m_{i2}, m_{i3}\right) = 0, \quad t = 1, ..., 4.$$

In order to ensure first-order Markov restrictions as in Assumption 2, we take the parameters $\rho_{1|2}$ and $\rho_{4|3}$ to be features of the covariance matrix of the ε's. Specifically, we take $\rho_{1|2}$ to be the population regression coefficient of ε_{i1} on ε_{i2} for workers who remain in the same firm in periods 1 and 2. Similarly, we take $\rho_{4|3}$ to be the regression coefficient of ε_{i4} on ε_{i3} for workers who remain in the same firm in periods 3 and 4. For simplicity, neither $\rho_{1|2}$ nor $\rho_{4|3}$ depend on the class of the firm, although this dependence may be allowed for (see below). Likewise, one could let the b_t's differ between stayers in (C7) and movers in (C8), see below.

The restrictions that $\rho_{1|2}$ and $\rho_{4|3}$ affect both the mean effects of firm classes on earnings for job movers and the covariance structure of earnings are consistent with Assumption 2. To see this in the case of $\rho_{4|3}$ (the argument for $\rho_{1|2}$ being similar), note that a mean independence counterpart to Assumption 2 (*ii*) is the following "backward" dynamic restriction:

$$\mathbb{E}\left(Y_{i4} \,|\, Y_{i1}, Y_{i2}, Y_{i3}, \alpha_i, k_{i2}, k_{i3}, m_{i1} = 0, m_{i2}, m_{i3} = 0\right) = \mathbb{E}\left(Y_{i4} \,|\, Y_{i3}, \alpha_i, k_{i3}, m_{i3} = 0\right),$$

which holds in model (C7)-(C8), for both movers and stayers (that is, whether $m_{i2} = 1$ or $m_{i2} = 0$), provided that:

$$\mathbb{E}\left(\varepsilon_{i4} \,|\, \varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \alpha_i, k_{i2}, k_{i3}, m_{i1} = 0, m_{i2}, m_{i3} = 0\right) = \rho_{4|3}\varepsilon_{i3}.$$

The structure of the dynamic model restricts how the effect of the previous firm class on log-earnings decays over time. Indeed, in (C8), log-earnings Y_{i3} after a job move may depend on the previous firm class k via the term $\xi_3(k)$. Log-earnings one period further apart from the move, Y_{i4}, still depend on k but the effect is $\rho_{4|3}\xi_3(k)$. In the special case where the ε's are uncorrelated, Y_{i4} does not depend on k, although Y_{i3} does. Analogously, as the probability of a job move between periods 2 and 3 (that is, that $m_{i2}=1$) depends on Y_{i2}, conditional on mobility log-earnings Y_{i1} and Y_{i2} before the move may depend on the class k' of the future firm. At the same time, the effect on first period's log-earnings is $\rho_{1|2}\xi_2(k')$, compared to $\xi_2(k')$ in period 2.

In addition, the model restricts how the effects of firm classes for job movers relate to those for job stayers. As an example, the effect of k' on Y_{i4} is a combination of the effect on Y_{i4} for job stayers ($a_4^s(k')$), and of the difference between the effects of k' on Y_{i3} for job movers and job stayers ($a_3^m(k')-a_3^s(k')$). In the absence of serial correlation in ε's this effect coincides with $a_4^s(k')$. In contrast, in the presence of serial correlation log-earnings of job movers and job stayers generally differ from each-other in all periods. This is again due to the fact that in this model mobility m_{it} depends on log-earnings Y_{it} directly.

From (C8) we have, for job movers between periods 2 and 3:

$$\begin{aligned} Y_{i1}-\rho_{1|2}Y_{i2} &= a_1^s(k)-\rho_{1|2}a_2^s(k)+\left[b_1(k)-\rho_{1|2}b_2(k)\right]\alpha_i+\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2}, \\ Y_{i4}-\rho_{4|3}Y_{i3} &= a_4^s(k')-\rho_{4|3}a_3^s(k')+\left[b_4(k')-\rho_{4|3}b_3(k')\right]\alpha_i+\varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3}. \end{aligned} \tag{C9}$$

Equation (C9) has a similar structure as the static model on two periods. As a result, one can derive moment restrictions analogous to (C6). For given $\rho_{1|2}$ and $\rho_{4|3}$, those restrictions are linear in parameters. It is therefore possible to adapt the results of the static model to identify the intercept and slope coefficients in (C9), as well as the means of α_i for job movers, under suitable rank conditions.

Specifically, we have the following result, where for simplicity we omit the conditioning on $m_{i1}=0$, $m_{i3}=0$, $k_{i1}=k_{i2}$, and $k_{i3}=k_{i4}$, all of which are true for both stayers (that is, $m_{i2}=0$) and movers ($m_{i2}=1$). For simplicity we abstract away from covariates X_{it}.

Theorem C2. *Suppose that $\rho_{1|2}$ and $\rho_{4|3}$ are known. Suppose also that the $b_t(k)$ coefficients are identical for job movers and job stayers (that is, that they are independent of m_{i2}).*

(i) Suppose that the conditions of Theorem C1 hold, with Y_{i1}, Y_{i2}, ε_{i1} and ε_{i2} being replaced by $Y_{i1}-\rho_{1|2}Y_{i2}$, $Y_{i4}-\rho_{4|3}Y_{i3}$, $\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2}$, and $\varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3}$, respectively. Then $a_1^s(k)-\rho_{1|2}a_2^s(k)$, $a_4^s(k')-\rho_{4|3}a_3^s(k')$, $b_1(k)-\rho_{1|2}b_2(k)$, $b_4(k')-\rho_{4|3}b_3(k')$, as well as $\mathbb{E}\left(\alpha_i \,|\, k_{i2},k_{i3},m_{i2}\right)$, $\mathrm{Var}\left(\alpha_i \,|\, k_{i2},k_{i3},m_{i2}\right)$, $\mathrm{Var}\left(\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2} \,|\, k_{i2}=k\right)$, and $\mathrm{Var}\left(\varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3} \,|\, k_{i3}=k\right)$, are all identified.

(ii) If, in addition to (i), the indicators $\mathbf{1}\{k_{i2}=k\}$, $\mathbf{1}\{k_{i3}=k'\}$, and the products $\mathbf{1}\{k_{i2}=k\}\times \mathbb{E}\left(\alpha_i \,|\, k_{i2},k_{i3},m_{i2}=1\right)$ are linearly independent conditional on $m_{i2}=1$, and the indicators $\mathbf{1}\{k_{i2}=k\}$, $\mathbf{1}\{k_{i3}=k'\}$, and the products $\mathbf{1}\{k_{i3}=k'\}\times\mathbb{E}\left(\alpha_i \,|\, k_{i2},k_{i3},m_{i2}=1\right)$ are linearly independent conditional on $m_{i2}=1$, then $a_2^m(k)$, $\xi_2^m(k)$, $b_2(k)$, $a_3^m(k)$, $\xi_3^m(k)$, and $b_3(k)$, are identified.

(iii) If, in addition to (i) and (ii), the $\mathbf{1}\{k_{i2}=k\}$ and $b_2(k_{i2})\times\mathbb{E}\left(\alpha_i \,|\, k_{i2},m_{i2}=0\right)$ are linearly independent conditional on $m_{i2}=0$, and the $\mathbf{1}\{k_{i3}=k\}$ and $b_3(k_{i3})\mathbb{E}\left(\alpha_i \,|\, k_{i3},m_{i2}=0\right)$ are linearly

independent conditional on $m_{i2}=0$, then $a_2^s(k)$ and $a_3^s(k)$ are identified.

(iv) If (i), (ii) and (iii) hold, then the covariance matrices of $\varepsilon_{i1},\varepsilon_{i2},\varepsilon_{i3},\varepsilon_{i4}$ are identified, for movers and stayers, conditional on every sequence of firm classes.

Proof. Part (i). This follows from Theorem C1.

Part (ii). This comes from:

$$\begin{aligned}
\mathbb{E}\left(Y_{i2}\,|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right) &= a_2^m(k)+\xi_2(k')+b_2(k)\mathbb{E}\left(\alpha_i\,|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right),\\
\mathbb{E}\left(Y_{i3}\,|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right) &= a_3^m(k')+\xi_3(k)+b_3(k')\mathbb{E}\left(\alpha_i\,|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right).
\end{aligned}$$

Part (iii). This comes from:

$$\begin{aligned}
\mathbb{E}\left(Y_{i2}\,|\,k_{i2}=k,m_{i2}=0\right) &= a_2^s(k)+b_2(k)\mathbb{E}\left(\alpha_i\,|\,k_{i2}=k,m_{i2}=0\right),\\
\mathbb{E}\left(Y_{i3}\,|\,k_{i2}=k,m_{i2}=0\right) &= a_3^s(k)+b_3(k)\mathbb{E}\left(\alpha_i\,|\,k_{i2}=k,m_{i2}=0\right).
\end{aligned}$$

Part (iv). For movers, we have:

$$\begin{aligned}
&\operatorname{Var}\left(\left(\begin{array}{c}\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2}\\ \varepsilon_{i2}\\ \varepsilon_{i3}\\ \varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3}\end{array}\right)\middle|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right)\\
&=\left(\begin{array}{cccc}\operatorname{Var}\left(\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2}|k_{i2}=k\right) & 0 & 0 & 0\\ 0 & V_{2kk'} & C_{23kk'} & 0\\ 0 & C_{23kk'} & V_{3kk'} & 0\\ 0 & 0 & 0 & \operatorname{Var}\left(\varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3}|k_{i3}=k'\right)\end{array}\right),
\end{aligned}$$

where $V_{2kk'}=\operatorname{Var}\left(\varepsilon_{i2}|k_{i2}=k,k_{i3}=k',m_{i2}=1\right)$, $C_{23kk'}=\operatorname{Cov}\left(\varepsilon_{i2},\varepsilon_{i3}|k_{i2}=k,k_{i3}=k',m_{i2}=1\right)$, and $V_{3kk'}=\operatorname{Var}\left(\varepsilon_{i3}|k_{i2}=k,k_{i3}=k',m_{i2}=1\right)$. Hence:

$$\begin{aligned}
&\operatorname{Var}\left(\left(\begin{array}{c}Y_{i1}-\rho_{1|2}Y_{i2}\\ Y_{i2}\\ Y_{i3}\\ Y_{i4}-\rho_{4|3}Y_{i3}\end{array}\right)\middle|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right)\\
&=\left(\begin{array}{c}b_1(k)-\rho_{1|2}b_2(k)\\ b_2(k)\\ b_3(k')\\ b_4(k')-\rho_{4|3}b_3(k')\end{array}\right)\times\operatorname{Var}\left(\alpha_i\,|\,k_{i2}=k,k_{i3}=k',m_{i2}=1\right)\times\left(\begin{array}{c}b_1(k)-\rho_{1|2}b_2(k)\\ b_2(k)\\ b_3(k')\\ b_4(k')-\rho_{4|3}b_3(k')\end{array}\right)'\\
&\quad+\left(\begin{array}{cccc}\operatorname{Var}\left(\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2}|k_{i2}=k\right) & 0 & 0 & 0\\ 0 & V_{2kk'} & C_{23kk'} & 0\\ 0 & C_{23kk'} & V_{3kk'} & 0\\ 0 & 0 & 0 & \operatorname{Var}\left(\varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3}|k_{i3}=k'\right)\end{array}\right),
\end{aligned}$$

from which we recover $V_{2kk'}$, $C_{23kk'}$ and $V_{2kk'}$. The variances $\mathrm{Var}\left(\varepsilon_{i1}|k_{i2}=k,k_{i3}=k',m_{i2}=1\right)$ and $\mathrm{Var}\left(\varepsilon_{i4}|k_{i2}=k,k_{i3}=k',m_{i2}=1\right)$ are then easy to recover. A similar argument allows recovering the covariance matrix of ε's for stayers.

∎

Parameters $\rho_{1|2}$ and $\rho_{4|3}$ may be recovered by exploiting the model's restrictions on the covariance structure of log-earnings. Below we explain how this can be done using restrictions on both job movers and job stayers. A simpler approach can be used under the additional assumptions that $b_t = b$ does not depend on t. Note that, while this condition imposes that interaction terms $b(k)\alpha_i$ do not vary over time within firm and worker, the effects of firm classes $a_t^s(k)$ and $a_t^m(k)$ are allowed to vary freely with time. Under this condition one can identify $\rho_{1|2}$ and $\rho_{4|3}$ using a set of covariance restrictions on job stayers alone. Indeed, within-firm log-earnings are the sum of a time-varying intercept ($a_t^s(k)$), a fixed effect ($b(k)\alpha_i$), and a first-order Markov shock (ε_{it}). The covariance matrix of the shocks and the variance of the fixed-effect are identified based on $T \geq 3$ periods, under suitable rank conditions. For example, in the model with four periods we have the following restrictions on $\rho_{1|2}$ and $\rho_{4|3}$:

$$\begin{aligned} \mathrm{Cov}\left(Y_{i1}-\rho_{1|2}Y_{i2}, Y_{i2}-Y_{i3}\,|\,k_{i1}=k_{i2}=k_{i3}=k_{i4}=k, m_{i1}=m_{i2}=m_{i3}=0\right)=0, \\ \mathrm{Cov}\left(Y_{i4}-\rho_{4|3}Y_{i3}, Y_{i3}-Y_{i2}\,|\,k_{i1}=k_{i2}=k_{i3}=k_{i4}=k, m_{i1}=m_{i2}=m_{i3}=0\right)=0. \end{aligned} \tag{C10}$$

These are familiar covariance restrictions in autoregressive models with fixed-effects. For example, the second equation in (C10) is the moment restriction corresponding to an instrumental variables regression of Y_{i4} on Y_{i3}, using $Y_{i3}-Y_{i2}$ as an instrument. A sufficient condition for identification of $\rho_{4|3}$ is thus that $\mathrm{Cov}\left(Y_{i3}, Y_{i3}-Y_{i2}\,|\,k_{i1}=k_{i2}=k_{i3}=k_{i4}=k, m_{i1}=m_{i2}=m_{i3}=0\right)\neq 0$. This condition requires that $\rho_{3|2}\neq 1$, where $\rho_{3|2}$ denotes the regression coefficient of ε_{i3} on ε_{i2}. Hence identification fails when ε_{it} follows exactly a unit root process. Finally, note that (C10) shows that one could easily allow for class-specific $\rho_{1|2}(k)$ and $\rho_{4|3}(k)$.

Dynamic model, unrestricted b's. Let us consider an extension of the dynamic interactive model where the b_t vary with t, and may differ between movers ($m_{i2}=1$) and stayers ($m_{i2}=0$). In order to enforce a Markovian structure as in Assumption 2 we impose:

$$b_1^s(k)-\rho_{1|2}b_2^s(k)=b_1^m(k)-\rho_{1|2}b_2^m(k), \quad b_4^s(k')-\rho_{4|3}b_3^s(k')=b_4^m(k')-\rho_{4|3}b_3^m(k').$$

Given the assumptions of Theorem C2, $b_1^m(k)-\rho_{1|2}b_2^m(k)$ and $b_4^m(k')-\rho_{4|3}b_3^m(k')$ are identified, together with $\mathbb{E}(\alpha_i \,|\, k_{i2}, k_{i3}, m_{i2})$. Moreover we have, for movers:

$$\begin{aligned}
\mathbb{E}\left(Y_{i1} \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right) &= a_1^s(k) + \rho_{1|2}(a_2^m(k) - a_2^s(k)) + \rho_{1|2}\xi_2(k') \\
&\quad + b_1^m(k)\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right), \\
\mathbb{E}\left(Y_{i2} \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right) &= a_2^m(k) + \xi_2(k') + b_2^m(k)\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right), \\
\mathbb{E}\left(Y_{i3} \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right) &= a_3^m(k') + \xi_3(k) + b_3^m(k')\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right), \\
\mathbb{E}\left(Y_{i4} \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right) &= a_4^s(k') + \rho_{4|3}(a_3^m(k') - a_3^s(k')) + \rho_{4|3}\xi_3(k) \\
&\quad + b_4^m(k')\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, k_{i3}=k', m_{i2}=1\right).
\end{aligned}$$

Hence, for given $\rho_{1|2}$ and $\rho_{4|3}$, the a's, b's, and ξ's are identified under a suitable rank condition as in Theorem C2 (ii).

For stayers we similarly have:

$$\begin{aligned}
\mathbb{E}\left(Y_{i1} \,|\, k_{i2}=k, m_{i2}=0\right) &= a_1^s(k) + b_1^s(k)\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, m_{i2}=0\right), \\
\mathbb{E}\left(Y_{i2} \,|\, k_{i2}=k, m_{i2}=0\right) &= a_2^s(k) + b_2^s(k)\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, m_{i2}=0\right), \\
\mathbb{E}\left(Y_{i3} \,|\, k_{i2}=k, m_{i2}=0\right) &= a_3^s(k) + b_3^s(k)\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, m_{i2}=0\right), \\
\mathbb{E}\left(Y_{i4} \,|\, k_{i2}=k, m_{i2}=0\right) &= a_4^s(k) + b_4^s(k)\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, m_{i2}=0\right).
\end{aligned}$$

Note that the means $\mathbb{E}(\alpha_i \,|\, k_{i2}=k, m_{i2}=0)$ are identified from Theorem C2. However, in this model with non-stationary and mobility-specific b's, the a_t^s and b_t^s are not identified based on mean restrictions alone.

Now, covariance restrictions on stayers imply:

$$\begin{aligned}
&\operatorname{Var}\left(\left(\begin{array}{c} Y_{i1} - \rho_{1|2}Y_{i2} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} - \rho_{4|3}Y_{i3} \end{array}\right) \middle|\, k_{i2}=k_{i3}=k, m_{i2}=0\right) \\
&= \left(\begin{array}{c} b_1^s(k) - \rho_{1|2}b_2^s(k) \\ b_2^s(k) \\ b_3^s(k) \\ b_4^s(k) - \rho_{4|3}b_3^s(k) \end{array}\right) \times \operatorname{Var}\left(\alpha_i \,|\, k_{i2}=k_{i3}=k, m_{i2}=0\right) \times \left(\begin{array}{c} b_1^s(k) - \rho_{1|2}b_2^s(k) \\ b_2^s(k) \\ b_3^s(k) \\ b_4^s(k) - \rho_{4|3}b_3^s(k) \end{array}\right)' \\
&\quad + \left(\begin{array}{cccc} \operatorname{Var}\left(\varepsilon_{i1} - \rho_{1|2}\varepsilon_{i2} | k_{i2}=k\right) & 0 & 0 & 0 \\ 0 & V_{2k}^s & C_{23k}^s & 0 \\ 0 & C_{23k}^s & V_{3k}^s & 0 \\ 0 & 0 & 0 & \operatorname{Var}\left(\varepsilon_{i4} - \rho_{4|3}\varepsilon_{i3} | k_{i3}=k\right) \end{array}\right),
\end{aligned} \tag{C11}$$

where: $V_{2k}^s = \operatorname{Var}(\varepsilon_{i2}|k_{i2}=k_{i3}=k, m_{i2}=0)$, $C_{23k}^s = \operatorname{Cov}(\varepsilon_{i2}, \varepsilon_{i3}|k_{i2}=k_{i3}=k, m_{i2}=0)$, and $V_{3k}^s = \operatorname{Var}(\varepsilon_{i3}|k_{i2}=k_{i3}=k, m_{i2}=0)$. Note that $b_1^s(k) - \rho_{1|2}b_2^s(k) = b_1^m(k) - \rho_{1|2}b_2^m(k)$, $b_4^s(k) - \rho_{4|3}b_3^s(k) =$

$b_4^m(k) - \rho_{4|3} b_3^m(k)$, $\mathrm{Var}\left(\varepsilon_{i1} - \rho_{1|2}\varepsilon_{i2} | k_{i2} = k\right)$, and $\mathrm{Var}\left(\varepsilon_{i4} - \rho_{4|3}\varepsilon_{i3} | k_{i3} = k\right)$ can be recovered from movers' mean and covariance restrictions. The system (C11) thus identifies $b_2^s(k)$, $b_3^s(k)$, V_{2k}^s, C_{23k}^s, V_{3k}^s, and $\mathrm{Var}\left(\alpha_i \,|\, k_{i2} = k_{i3} = k, m_{i2} = 0\right)$, under suitable rank conditions.

Lastly, all the arguments above have been conducted for known $\rho_{1|2}$ and $\rho_{4|3}$. The ρ's may be recovered from jointly imposing covariance restrictions for stayers in (C11), and for movers in the following system:

$$\mathrm{Var}\left(\left(\begin{array}{c} Y_{i1} - \rho_{1|2} Y_{i2} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} - \rho_{4|3} Y_{i3} \end{array}\right) \middle| k_{i2} = k, k_{i3} = k', m_{i2} = 1\right)$$

$$= \left(\begin{array}{c} b_1^m(k) - \rho_{1|2} b_2^m(k) \\ b_2^m(k) \\ b_3^m(k') \\ b_4^m(k') - \rho_{4|3} b_3^m(k') \end{array}\right) \times \mathrm{Var}\left(\alpha_i \,|\, k_{i2} = k, k_{i3} = k', m_{i2} = 1\right) \times \left(\begin{array}{c} b_1^m(k) - \rho_{1|2} b_2^m(k) \\ b_2^m(k) \\ b_3^m(k') \\ b_4^m(k') - \rho_{4|3} b_3^m(k') \end{array}\right)'$$

$$+ \left(\begin{array}{cccc} \mathrm{Var}\left(\varepsilon_{i1} - \rho_{1|2}\varepsilon_{i2} | k_{i2} = k\right) & 0 & 0 & 0 \\ 0 & V_{2kk'} & C_{23kk'} & 0 \\ 0 & C_{23kk'} & V_{3kk'} & 0 \\ 0 & 0 & 0 & \mathrm{Var}\left(\varepsilon_{i4} - \rho_{4|3}\varepsilon_{i3} | k_{i3} = k'\right) \end{array}\right),$$

across all values of k, k'. In this case also, identification relies on a rank condition to be satisfied.

C.2 Estimation algorithms in interactive models

Consider the static interactive model (1) on two periods. The mean restrictions in (C6) being linear in parameters, estimation can be based on linear IV techniques. The LIML estimator is particularly convenient here, as it is invariant to scaling of the moment conditions. In practice, this means that the normalization on intercept and slope parameters (e.g., $a_1(1) = 0$, $b_1(1) = 1$) is immaterial for the results. In addition, LIML is computationally convenient as it is the solution to a minimum eigenvalue problem.[29]

The identification results in Theorem C1 thus suggest the following multi-step estimation method. First, estimate firm classes $\widehat{k}(j)$. Given the estimated firm classes, construct the instruments Z_i in (C6), and estimate intercepts, slopes, and coefficients associated with observables using LIML, see

[29]Specifically, let us write the moment restrictions in (C6) as $\mathbb{E}(Z_i' W_i \beta) = 0$. Then the LIML estimator is given, up to scale, by:

$$\widehat{\beta} = \underset{b}{\mathrm{argmin}} \, \frac{b' W' Z (Z'Z)^{-1} Z' W b}{b' W' W b}.$$

Alternatively, $\widehat{\beta}$ is the minimum eigenvalue of the matrix $(W'W)^{-1} W' Z (Z'Z)^{-1} Z' W$.

part (i) in Theorem C1. Then estimate means of α_i using linear regression, see also part (i). Finally, estimate variances of α_i, ε_{i1}, and ε_{i2} using empirical counterparts to the covariance restrictions in parts (ii) and (iii) in Theorem C1. The latter restrictions are also linear in parameters, so computation is straightforward.[30]

We now describe the estimation algorithms in static and dynamic interactive models.

Static case. We consider estimation in the static interactive model (1) on two periods. The algorithm is as follows.

1. Estimate firm classes $\widehat{k}(j)$.

2. Perform the following sub-steps:

   - Construct $\widehat{Z}_i$ from dummies $\widehat{k}_{i1}$ and $\widehat{k}_{i2}$ and their interactions, as well as interactions with (X_{i1}, X_{i2}). Estimate parameters $\widehat{\widetilde{a}}_t(k)$, $\widehat{\tau}_t(k)$, and $\widehat{\widetilde{c}}_t(k)$ using LIML based on (C6) with scale and location normalizations.[31] Recover $\widehat{b}_t(k) = 1/\widehat{\tau}_t(k)$, $\widehat{a}_t(k) = \widehat{b}_t(k)\widehat{\widetilde{a}}_t(k)$, and $\widehat{c}_t$ as a mean of the $\widehat{\widetilde{c}}_t(k) = \widehat{b}_t(k)\widehat{\widetilde{c}}_t(k)$ over k, weighted by the probabilities $\Pr(\widehat{k}_{i1} = k)$.

   - Let $\mu^m_{kk'} = \mathbb{E}(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1)$, and $\mu_k = \mathbb{E}(\alpha_i \,|\, k_{i1} = k)$. Estimate $\widehat{\mu}^m_{kk'}$ as the mean of:
$$\frac{1}{2}\sum_{t=1}^{2} \widehat{\tau}_t(\widehat{k}_{it})Y_{it} - \widehat{\widetilde{a}}_t(\widehat{k}_{it}) - X'_{it}\widehat{\widetilde{c}}_t(\widehat{k}_{it}) \tag{C12}$$
given $\widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1$. Estimate $\widehat{\mu}_k$ as the mean of (C12) given $\widehat{k}_{i1} = k$. Note that it is easy to also recover estimates of means of α_i for job stayers (that is, $m_{i1} = 0$). Construct $\widetilde{Y}_{it} = Y_{it} - X'_{it}\widehat{c}_t$.

   - Estimate the variances $\mathrm{Var}\,(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1) = v^m_{kk'}$, $\mathrm{Var}\,(\varepsilon_{i1} \,|\, k_{i1} = k) = V_{1k}$, and $\mathrm{Var}\,(\varepsilon_{i2} \,|\, k_{i2} = k) = V_{4k}$ by minimizing:
$$\sum_{k=1}^{K}\sum_{k'=1}^{K} N^m_{kk'} \left\| \begin{pmatrix} \widehat{\mathrm{Var}}\left(\widetilde{Y}_{i1} \,|\, \widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1\right) \\ \widehat{\mathrm{Cov}}\left(\widetilde{Y}_{i1}, \widetilde{Y}_{i2} \,|\, \widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1\right) \\ \widehat{\mathrm{Var}}\left(\widetilde{Y}_{i2} \,|\, \widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1\right) \end{pmatrix} - \begin{pmatrix} \widehat{b}_1(k)^2 v^m_{kk'} + V_{1k} \\ \widehat{b}_1(k)\widehat{b}_2(k') v^m_{kk'} \\ \widehat{b}_2(k')^2 v^m_{kk'} + V_{2k'} \end{pmatrix} \right\|^2,$$
subject to all $v^m_{kk'} \geq 0$, $V_{1k} \geq 0$, $V_{2k'} \geq 0$, where $\widehat{\mathrm{Var}}$ and $\widehat{\mathrm{Cov}}$ denote empirical variances and covariances, respectively, and $N^m_{kk'}$ denotes the number of job movers from $\widehat{k}_{i1} = k$ to

[30] In practice, it may be useful to explicitly impose that variances be non-negative when fitting covariance restrictions. This requires solving quadratic programming problems, which are convex and numerically well-behaved, see below.

[31] Note that an additive specification is obtained as a special case, when one imposes that $b_t(k) = 1$ for all t, k in this step.

$\widehat{k}_{i2} = k'$. Lastly, estimate $\mathrm{Var}\left(\alpha_i \,|\, k_{i1} = k\right) = v_k$ by minimizing:

$$\sum_{k=1}^{K} N_k \left|\left| \widehat{\mathrm{Var}}\left(\widetilde{Y}_{i1} \,|\, \widehat{k}_{i1} = k\right) - \widehat{b}_1(k)^2 v_k - \widehat{V}_{1k} \right|\right|^2,$$

subject to all $v_k \geq 0$, where N_k denotes the number of workers in firm class $\widehat{k}_{i1} = k$ in period 1.

Dynamic case. We consider estimation in the dynamic interactive model on four periods (C7)-(C8). We focus on the case where the b coefficients are stationary and common across movers and stayers. A more general estimation algorithm is readily constructed. For simplicity we do not include covariates X_{it}, although their coefficients can be easily estimated from the LIML sub-step. The algorithm is as follows.

1. Estimate firm classes $\widehat{k}(j)$.

2. Perform the following sub-steps:

   - Consider the following objective function:

$$Q(\rho_1, \rho_4) = \min \sum_{k=1}^{K} N_k^s \left|\left| \begin{pmatrix} \widehat{\mathrm{Var}}\left(Y_{i1} - \rho_1 Y_{i2} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\mathrm{Cov}}\left(Y_{i1} - \rho_1 Y_{i2}, Y_{i2} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\mathrm{Cov}}\left(Y_{i1} - \rho_1 Y_{i2}, Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\mathrm{Cov}}\left(Y_{i1} - \rho_1 Y_{i2}, Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\mathrm{Cov}}\left(Y_{i2}, Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\mathrm{Cov}}\left(Y_{i3}, Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\mathrm{Var}}\left(Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \end{pmatrix} - \begin{pmatrix} (1-\rho_1)^2 \widetilde{v}_k^s + V_{1k} \\ (1-\rho_1)\widetilde{v}_k^s \\ (1-\rho_1)\widetilde{v}_k^s \\ (1-\rho_1)(1-\rho_4)\widetilde{v}_k^s \\ (1-\rho_4)\widetilde{v}_k^s \\ (1-\rho_4)\widetilde{v}_k^s \\ (1-\rho_4)^2 \widetilde{v}_k^s + V_{4k} \end{pmatrix} \right|\right|^2,$$

   subject to all $\widetilde{v}_k^s \geq 0$, $V_{1k} \geq 0$, $V_{4k} \geq 0$. Estimate $\widehat{\rho}_{1|2}$ and $\widehat{\rho}_{4|3}$ as:

$$\left(\widehat{\rho}_{1|2}, \widehat{\rho}_{4|3}\right) = \underset{(\rho_1, \rho_4)}{\mathrm{argmin}}\; Q(\rho_1, \rho_4).$$

   In practice, different minimum-distance weights can be used.

- Let $c_1(k) = a_1^s(k) - \rho_{1|2}a_2^s(k)$, $c_4(k) = a_4^s(k) - \rho_{4|3}a_3^s(k)$, $d_1(k) = b(k) - \rho_{1|2}b(k)$, and $d_4(k) = b(k) - \rho_{4|3}b(k)$. Construct $\widehat{Z}_i$ from dummies $\widehat{k}_{i2}$ and $\widehat{k}_{i3}$ and their interactions. Estimate parameters $\widehat{\tau}_1(k) = 1/\widehat{d}_1(k)$, $\widehat{\tau}_4(k) = 1/\widehat{d}_4(k)$, $\widehat{\widetilde{c}}_1(k) = \widehat{c}_1(k)/\widehat{d}_1(k)$, and $\widehat{\widetilde{c}}_4(k) = \widehat{c}_4(k)/\widehat{d}_4(k)$ using LIML based on:

$$\mathbb{E}\left[Z_i\left(\tau_4(k_{i3})\left(Y_{i4} - \rho_{4|3}Y_{i3}\right) - \tau_1(k_{i2})\left(Y_{i1} - \rho_{1|2}Y_{i2}\right) - \widetilde{c}_4(k_{i3}) + \widetilde{c}_1(k_{i2})\right) \mid m_{i2} = 1\right] = 0,$$

imposing scale and location normalizations and replacing $\rho_{1|2}$ and $\rho_{4|3}$ by $\widehat{\rho}_{1|2}$ and $\widehat{\rho}_{4|3}$. This yields estimates $\widehat{d}_1(k)$, $\widehat{d}_4(k)$, $\widehat{c}_1(k)$, and $\widehat{c}_4(k)$. This also yields estimates of

$$\widehat{b}(k) = \frac{\widehat{d}_1(k)}{2(1-\widehat{\rho}_{1|2})} + \frac{\widehat{d}_4(k)}{2(1-\widehat{\rho}_{4|3})}.$$

- Let $\mu_{kk'}^m = \mathbb{E}(\alpha_i \mid k_{i2} = k, k_{i3} = k', m_{i2} = 1)$. Estimate $\widehat{\mu}_{kk'}^m$ as the mean of:

$$\frac{1}{2}\left(\widehat{\tau}_1(\widehat{k}_{i2})\left(Y_{i1} - \widehat{\rho}_{1|2}Y_{i2}\right) - \widehat{\widetilde{a}}_1(\widehat{k}_{i2}) + \widehat{\tau}_4(\widehat{k}_{i3})\left(Y_{i4} - \widehat{\rho}_{4|3}Y_{i3}\right) - \widehat{\widetilde{a}}_4(\widehat{k}_{i3})\right) \tag{C13}$$

given $\widehat{k}_{i2} = k, \widehat{k}_{i3} = k', m_{i2} = 1$. Let $\mu_k^s = \mathbb{E}(\alpha_i \mid k_{i2} = k_{i3} = k, m_{i2} = 0)$. Estimate $\widehat{\mu}_k^s$ as the mean of (C13) given $\widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0$.

- Estimate $\widehat{a}_t^s(k)$ as the mean of:

$$Y_{it} - \widehat{b}(k)\widehat{\mu}_k^s,$$

given $\widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0$.

- Estimate $\widehat{a}_2^m(k)$, $\widehat{a}_3^m(k')$, $\widehat{\xi}_2(k)$, and $\widehat{\xi}_3(k')$ by minimizing:

$$\sum_{k=1}^{K}\sum_{k'=1}^{K} N_{kk'}^m \left\| \begin{pmatrix} \widehat{\mathbb{E}}\left(Y_{i1} \mid \widehat{k}_{i2} = k, \widehat{k}_{i3} = k', m_{i2} = 1\right) + \widehat{\rho}_{1|2}\widehat{a}_2^s(k) - \widehat{a}_1^s(k) - \widehat{b}(k)\widehat{\mu}_{kk'}^m \\ \widehat{\mathbb{E}}\left(Y_{i2} \mid \widehat{k}_{i2} = k, \widehat{k}_{i3} = k', m_{i2} = 1\right) - \widehat{b}(k)\widehat{\mu}_{kk'}^m \\ \widehat{\mathbb{E}}\left(Y_{i3} \mid \widehat{k}_{i2} = k, \widehat{k}_{i3} = k', m_{i2} = 1\right) - \widehat{b}(k')\widehat{\mu}_{kk'}^m \\ \widehat{\mathbb{E}}\left(Y_{i4} \mid \widehat{k}_{i2} = k, \widehat{k}_{i3} = k', m_{i2} = 1\right) + \widehat{\rho}_{4|3}\widehat{a}_3^s(k') - \widehat{a}_4^s(k') - \widehat{b}(k')\widehat{\mu}_{kk'}^m \end{pmatrix} - \begin{pmatrix} \widehat{\rho}_{1|2}\left(a_2^m(k) + \xi_2(k')\right) \\ a_2^m(k) + \xi_2(k') \\ a_3^m(k') + \xi_3(k) \\ \widehat{\rho}_{4|3}\left(a_3^m(k') + \xi_3(k)\right) \end{pmatrix} \right\|^2,$$

which is a quadratic objective function.

- Let $V_{kk'}^{\alpha m} = \mathrm{Var}\left(\alpha_i \mid k_{i2} = k, k_{i3} = k', m_{i2} = 1\right)$, $V_{2kk'} = \mathrm{Var}\left(\varepsilon_{i2} | k_{i2} = k, k_{i3} = k', m_{i2} = 1\right)$, $C_{23kk'} = \mathrm{Cov}\left(\varepsilon_{i2}, \varepsilon_{i3} | k_{i2} = k, k_{i3} = k', m_{i2} = 1\right)$, $V_{3kk'} = \mathrm{Var}\left(\varepsilon_{i3} | k_{i2} = k, k_{i3} = k', m_{i2} = 1\right)$, $V_k^{\varepsilon 12} = \mathrm{Var}\left(\varepsilon_{i1} - \rho_{1|2}\varepsilon_{i2} | k_{i2} = k\right)$, and $V_k^{\varepsilon 43} = \mathrm{Var}\left(\varepsilon_{i4} - \rho_{4|3}\varepsilon_{i3} | k_{i3} = k\right)$. Estimate $\widehat{V}_{kk'}^{\alpha m}$,

$\widehat{V}_{2kk'}$, $\widehat{C}_{23kk'}$, $\widehat{V}_{3kk'}$, $\widehat{V}_{k}^{\varepsilon_{12}}$, and $\widehat{V}_{k}^{\varepsilon_{43}}$ by minimizing:

$$\sum_{k=1}^{K}\sum_{k'=1}^{K} N_{kk'}^{m}\left\|\widehat{\text{Var}}\left(\left(\begin{array}{c} Y_{i1}-\widehat{\rho}_{1|2}Y_{i2} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4}-\widehat{\rho}_{4|3}Y_{i3} \end{array}\right)\middle|\,\widehat{k}_{i2}=k,\widehat{k}_{i3}=k',m_{i2}=1\right)\right.$$
$$-\left(\begin{array}{c} \widehat{b}(k)-\widehat{\rho}_{1|2}\widehat{b}(k) \\ \widehat{b}(k) \\ \widehat{b}(k') \\ \widehat{b}(k')-\widehat{\rho}_{4|3}\widehat{b}(k') \end{array}\right)\times V_{kk'}^{\alpha m}\times\left(\begin{array}{c} \widehat{b}(k)-\widehat{\rho}_{1|2}\widehat{b}(k) \\ \widehat{b}(k) \\ \widehat{b}(k') \\ \widehat{b}(k')-\widehat{\rho}_{4|3}\widehat{b}(k') \end{array}\right)'$$
$$\left.-\left(\begin{array}{cccc} V_{k}^{\varepsilon_{12}} & 0 & 0 & 0 \\ 0 & V_{2kk'} & C_{23kk'} & 0 \\ 0 & C_{23kk'} & V_{3kk'} & 0 \\ 0 & 0 & 0 & V_{k'}^{\varepsilon_{43}} \end{array}\right)\right\|^{2},$$

subject to all $V_{kk'}^{\alpha m}\geq 0$, $V_{k}^{\varepsilon_{12}}\geq 0$, $V_{k'}^{\varepsilon_{43}}\geq 0$, $V_{2kk'}\geq 0$, $V_{3kk'}\geq 0$. This is a quadratic programming problem. In addition one may impose the quadratic constraint: $C_{23kk'}^{2}\leq V_{2kk'}V_{3kk'}$. If needed, estimate $\text{Var}\,(\varepsilon_{i1}|k_{i2}=k,k_{i3}=k',m_{i2}=1)$, $\text{Cov}(\varepsilon_{i1},\varepsilon_{i2}|k_{i2}{=}k,k_{i3}{=}k',m_{i2}=1)$, $\text{Var}\,(\varepsilon_{i4}|k_{i2}=k,k_{i3}=k',m_{i2}=1)$, and $\text{Cov}\,(\varepsilon_{i3},\varepsilon_{i4}|k_{i2}=k,k_{i3}=k',m_{i2}=1)$ by simple matrix inversion.

- Let $V_{k}^{\alpha s}=\text{Var}\,(\alpha_i\,|\,k_{i2}=k_{i3}=k,m_{i2}=0)$, $V_{2k}^{s}=\text{Var}\,(\varepsilon_{i2}|k_{i2}=k_{i3}=k,m_{i2}=0)$, $C_{23k}^{s}=\text{Cov}\,(\varepsilon_{i2},\varepsilon_{i3}|k_{i2}=k_{i3}=k,m_{i2}=0)$, and $V_{3k}^{s}=\text{Var}\,(\varepsilon_{i3}|k_{i2}=k_{i3}=k,m_{i2}=0)$. Estimate $\widehat{V}_{k}^{\alpha s}$, $\widehat{V}_{2k}^{s}$, $\widehat{C}_{23k}^{s}$, and $\widehat{V}_{3k}^{s}$ by minimizing:

$$\sum_{k=1}^{K} N_{k}^{s}\left\|\widehat{\text{Var}}\left(\left(\begin{array}{c} Y_{i1}-\widehat{\rho}_{1|2}Y_{i2} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4}-\widehat{\rho}_{4|3}Y_{i3} \end{array}\right)\middle|\,\widehat{k}_{i2}=\widehat{k}_{i3}=k,m_{i2}=0\right)\right.$$
$$-\left(\begin{array}{c} \widehat{b}(k)-\widehat{\rho}_{1|2}\widehat{b}(k) \\ \widehat{b}(k) \\ \widehat{b}(k) \\ \widehat{b}(k)-\widehat{\rho}_{4|3}\widehat{b}(k) \end{array}\right)\times V_{kk}^{\alpha s}\times\left(\begin{array}{c} \widehat{b}(k)-\widehat{\rho}_{1|2}\widehat{b}(k) \\ \widehat{b}(k) \\ \widehat{b}(k) \\ \widehat{b}(k)-\widehat{\rho}_{4|3}\widehat{b}(k) \end{array}\right)'$$
$$\left.-\left(\begin{array}{cccc} \widehat{V}_{k}^{\varepsilon_{12}} & 0 & 0 & 0 \\ 0 & V_{2k}^{s} & C_{23k}^{s} & 0 \\ 0 & C_{23k}^{s} & V_{3k}^{s} & 0 \\ 0 & 0 & 0 & \widehat{V}_{k}^{\varepsilon_{43}} \end{array}\right)\right\|^{2},$$

subject to all $V_k^{\alpha s} \geq 0$, $V_{2k}^s \geq 0$, $C_{23k}^s \geq 0$, and $V_{3k}^s \geq 0$. This is another quadratic programming problem. Here also one may impose as quadratic constraints: $(C_{23k}^s)^2 \leq V_{2k}^s V_{3k}^s$. If needed, estimate $\text{Var}\,(\varepsilon_{i1}|k_{i2} = k_{i3} = k, m_{i2} = 0)$, $\text{Cov}\,(\varepsilon_{i1}, \varepsilon_{i2}|k_{i2}{=}k_{i3}{=}k, m_{i2}{=}0)$, $\text{Var}(\varepsilon_{i4}| k_{i2}{=}k_{i3}{=}k, m_{i2}{=}0)$, and $\text{Cov}\,(\varepsilon_{i3}, \varepsilon_{i4} \,|\, k_{i2}{=}k_{i3}{=}k, m_{i2}{=}0)$.

C.3 Interactive models on T periods

Here we consider the dynamic interactive model on T periods. The static interactive model is a special case of the latter. An important feature of interactive models is that they are defined conditionally on a sequence of firm classes and mobility choices. We thus start by deriving restrictions implied by Assumption 2 on earnings distributions conditional on the entire sequences of k_{it} and m_{it}. Given that we work with interactive regression models, we emphasize the implications of Assumption 2 on means and variances. We assume strictly exogenous X's for simplicity, and focus on models where the b_t's do not depend on mobility (although the a's do).

The first-order Markov structure implies the following "forward" and "backward" restrictions, denoting $Z_i^{t:t+s} = (Z_{it}, ..., Z_{i,t+s})$.

- Forward restrictions:

$$\mathbb{E}\left[Y_{it} \,|\, Y_{i,t+s}, \alpha_i, k_i^T, m_i^{T-1}, X_i^T\right] = \mathbb{E}\left[Y_{it} \,|\, Y_{i,t+s}, \alpha_i, k_i^{1:t+s}, m_i^{1:t+s-1}, X_i^T\right], \quad s > 0.$$

- Backward restrictions:

$$\mathbb{E}\left[Y_{it} \,|\, Y_{i,t-s}, \alpha_i, k_i^T, m_i^{T-1}, X_i^T\right] = \mathbb{E}\left[Y_{it} \,|\, Y_{i,t-s}, \alpha_i, k_i^{t-s:T}, m_i^{t-s:T-1}, X_i^T\right], \quad s > 0.$$

A simple regression model that satisfies these conditions is defined as follows, *conditionally* on a sequence $(k_i^T, m_i^{T-1}, X_i^T)$:

$$\begin{aligned} Y_{it} &= a_{tt}(k_{it}, m_{i,t-1}, m_{it}) + b_t(k_{it})\alpha_i + X_{it}'c_t + \varepsilon_{it} \\ &\quad + \sum_{s=1}^{T-t} \left(\rho_{t|t+s} a_{t+s,t+s}(k_{i,t+s}, m_{i,t+s-1}, m_{i,t+s}) + \rho_{t|t+s-1}\xi_{t+s}^f(k_{i,t+s}, m_{i,t+s-1})\right) \\ &\quad + \sum_{s=1}^{t-1} \left(\rho_{t|t-s} a_{t-s,t-s}(k_{i,t-s}, m_{i,t-s-1}, m_{i,t-s}) + \rho_{t|t-s-1}\xi_{t-s}^b(k_{i,t-s}, m_{i,t-s})\right), \end{aligned}$$

where ε_{it} is first-order Markov with $\mathbb{E}\left(\varepsilon_{it} \,|\, \alpha_i, k_i^T, m_i^{T-1}, X_i^T\right) = 0$, and, for all $s > 0$:

$$\mathbb{E}\left(\varepsilon_{it} \,|\, \varepsilon_i^{1:t-s}, \alpha_i, k_i^T, m_i^{T-1}, X_i^T\right) = \rho_{t|t-s}\varepsilon_{i,t-s}, \quad \mathbb{E}\left(\varepsilon_{it} \,|\, \varepsilon_i^{t+s:T}, \alpha_i, k_i^T, m_i^{T-1}, X_i^T\right) = \rho_{t|t+s}\varepsilon_{i,t+s}.$$

As a result: $\rho_{t+s+m|t} = \rho_{t+s+m|t+s}\rho_{t+s|t}$ and $\rho_{t|t+s+m} = \rho_{t|t+s}\rho_{t+s|t+s+m}$ for all $s > 0, m > 0$.

Estimation. The main difference with the estimation of the dynamic model on four periods is in the estimation of the mean parameters, i.e. the a's, b's, c's, and ξ's given the ρ's. Let $\tau_t(k) = 1/b_t(k)$, and let:

$$\begin{aligned} W_{it}'\gamma_t \quad &= \quad a_{tt}(k_{it}, m_{i,t-1}, m_{it}) + X_{it}'c_t \\ &+ \sum_{s=1}^{T-t} \left(\rho_{t|t+s} a_{t+s,t+s}(k_{i,t+s}, m_{i,t+s-1}, m_{i,t+s}) + \rho_{t|t+s-1} \xi_{t+s}^{f}(k_{i,t+s}, m_{i,t+s-1}) \right) \\ &+ \sum_{s=1}^{t-1} \left(\rho_{t|t-s} a_{t-s,t-s}(k_{i,t-s}, m_{i,t-s-1}, m_{i,t-s}) + \rho_{t|t-s-1} \xi_{t-s}^{b}(k_{i,t-s}, m_{i,t-s}) \right). \end{aligned}$$

We have:

$$\mathbb{E}\left[\tau_t(k_{it}) \left(Y_{it} - W_{it}'\gamma_t \right) \mid k_i^T, m_i^{T-1}, X_i^T \right] = 0,$$

which is a quadratic conditional moment restriction. Letting $Z_{it} = Z_{it}(k_i^T, m_i^{T-1}, X_i^T)$ be a vector of instruments we can base estimation on:

$$\mathbb{E}\left[Z_{it}\tau_t(k_{it}) \left(Y_{it} - W_{it}'\gamma_t \right) \right] = 0.$$

In order to ensure invariance to normalization, one can solve a continuously updated GMM problem such as:

$$\min_{(\tau,\gamma)} \frac{\sum_{i=1}^N \sum_{t=1}^T Z_{it}'\tau_t(k_{it}) \left(Y_{it} - W_{it}'\gamma_t \right) \left(\sum_{i=1}^N \sum_{t=1}^T Z_{it}Z_{it}' \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T Z_{it}\tau_t(k_{it}) \left(Y_{it} - W_{it}'\gamma_t \right)}{\sum_{i=1}^N \sum_{t=1}^T \tau_t(k_{it}) \left(Y_{it} - W_{it}'\gamma_t \right)^2}.$$

In practice this objective function may be minimized iteratively, iterating between τ's and γ's, each step corresponding to a LIML-like minimum eigenvalue problem.

Finally, for implementation it is important to choose a parsimonious set of instruments.

D Estimation on data from 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:

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

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

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

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

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

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.

Wages. We then derive the wage function 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 ($q_k(\alpha)$ in the body of the paper):

$$q_y(x) = \frac{h(x,y)}{1 - v(y)},$$

and the type proportions for job movers ($p_{k'k}(\alpha)$ in the main text), which are given by:

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

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 ($F_{k\alpha}$ 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:

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

These probabilities are symmetric in (y, y'). In the context of Theorem 1 this means that Assumption 3 *i*) is not satisfied, as $a(\alpha) = 1$ for all α. 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 E3 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:

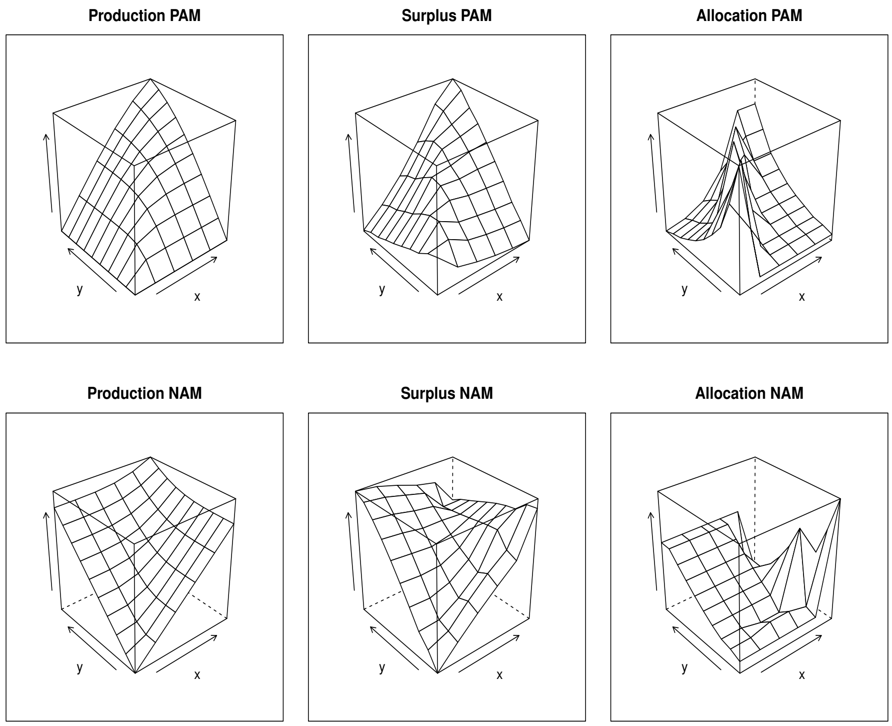
$$f(x,y) = a + \left(\nu x^\rho + (1-\nu)y^\rho\right)^{1/\rho},$$

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

Figures D1 and D2 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.

In Table D1 we report the results of variance decompositions based on the data generated according to the model, and based on estimates from our static finite mixture model based on those data. We use $L = 6$ (worker types) and $K = 10$ (firm classes) in estimation, and consider two scenarios for L and K 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

Figure D1: Model solutions: production, surplus and allocation



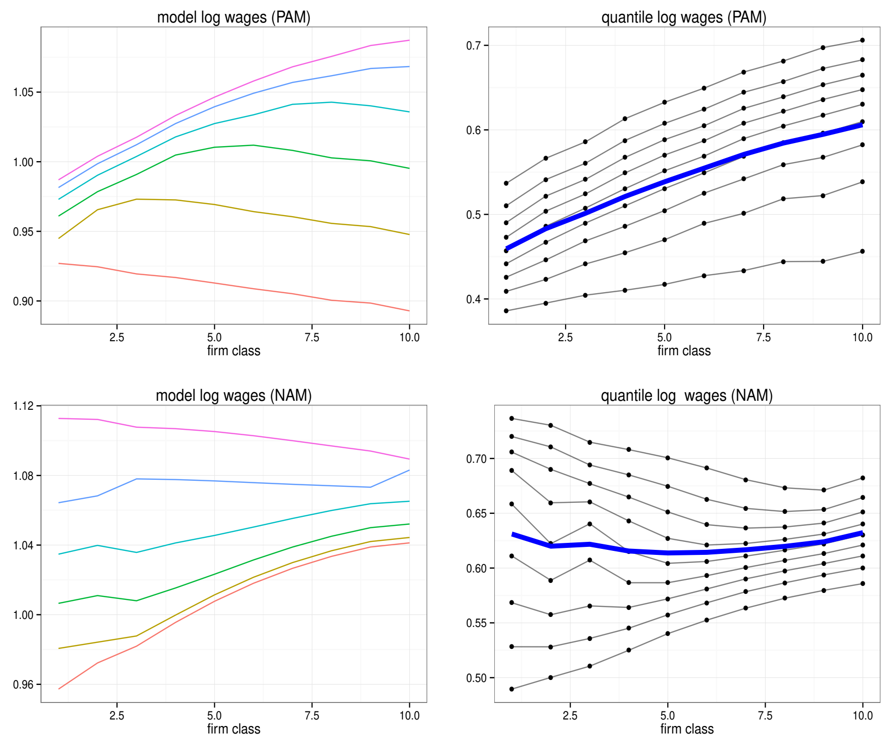
Notes: The graphs show the model solution in terms of production $f(x, y)$, surplus $S(x, y)$, and allocation $h(x, y)$. Positive assortative matching (top panel), and negative assortative matching (bottom panel).

Table D1: 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 $L \times K$, where L is the number of worker types and K 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.

Figure D2: 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).

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 D1 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 components rather well, even when the number of types and classes 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.

E Additional empirical results

Table E2: Data description, larger sample

	all		continuing firms, full year employed			
years:	2002-2004		2002-2004		2001-2005	
	all	movers	all	movers	all	movers
number of workers	795,419	88,771	599,775	19,557	442,757	9,645
number of firms	50,448	17,887	43,826	7,557	36,928	4,248
number of firms ≥ 10	26,834	13,233	23,389	6,231	20,557	3,644
number of firms ≥ 50	4,876	3,974	4,338	2,563	3,951	1,757
mean firm reported size	36.41	76.5	37.59	132.33	39.67	184.77
median firm reported size	10	18	10	28	11	36
firm reported size for median worker	154	158	154	159	162	176
firm actual size for median worker	83	23	72	5	64	3
% high school drop out	19.6%	15%	20.6%	14%	21.5%	14.7%
% high school graduates	56.6%	56.7%	56.7%	57.3%	57%	59%
% some college	23.7%	28.3%	22.7%	28.7%	21.4%	26.3%
% workers younger than 30	19.3%	26.8%	16.8%	28%	13.9%	23.8%
% workers between 31 and 50	56.8%	56.7%	57.2%	59%	59.4%	62.1%
% workers older than 51	23.9%	16.5%	26%	13%	26.7%	14.2%
% workers in manufacturing	43.5%	35.4%	45.4%	35.1%	48.5%	40.4%
% workers in services	27%	34.3%	25.3%	33.7%	22.4%	27.8%
% workers in retail and trade	16.2%	15%	16.7%	20.3%	16.3%	20.8%
% workers in construction	13.3%	15.3%	12.6%	10.9%	12.8%	11%
mean log-earnings	10.16	10.15	10.18	10.17	10.19	10.17
variance of log-earnings	0.146	0.2	0.124	0.166	0.113	0.148
between-firm variance of log-earnings	0.055	0.104	0.0475	0.1026	0.0441	0.0947
mean log-value-added per worker	15.28	15.86	15.3	16.35	15.37	16.63

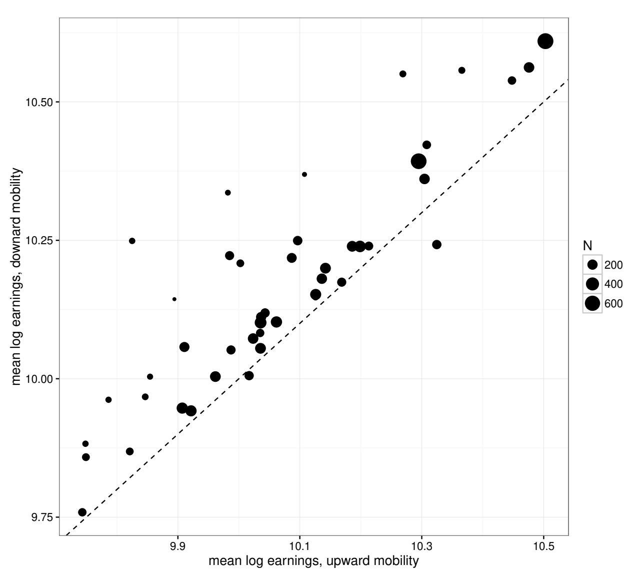
Notes: Swedish registry data. Males, employed in the last quarter of 2002 and the first quarter of 2004. Figures for 2002.

Table E3: Number of job movers between firm classes

	firm class in period 2									
firm class in period 1	1	2	3	4	5	6	7	8	9	10
1	80	109	45	88	86	43	54	40	45	18
2	118	534	255	260	229	143	191	140	85	34
3	58	201	207	150	233	217	107	149	60	23
4	94	277	198	426	299	197	264	181	157	120
5	46	249	193	247	309	248	211	234	131	59
6	29	95	128	125	271	253	142	284	152	61
7	39	141	108	340	289	162	292	232	215	112
8	33	250	113	193	255	296	270	434	654	217
9	12	75	73	149	226	284	247	562	685	671
10	8	31	27	115	65	63	155	208	656	918

Notes: Males, fully employed in the same firm in 2002 and in 2004, continuously existing firms. Movers from firm class k (vertical axis) to firm class k′ (horizontal axis).

Figure E3: Earnings of job movers



Notes: Males, fully employed in the same firm in 2002 and in 2004, continuously existing firms. Mean log-earnings over 2002 and 2004 of movers from firm class k to firm class k' (x-axis), and of movers from k' to k (y-axis), where $k < k'$.